Comparing forecasts of Latvia's GDP using simple seasonal ARIMA models and direct versus indirect approach

Ginters Bušs

Central Statistical Bureau of Latvia, Riga Technical University, Faculty of Computer Science and Information Technology, Department of Probability Theory and Mathematical Statistics

6 August 2009

Comparing forecasts of Latvia's GDP using Simple Seasonal ARIMA models and Direct versus Indirect Approach

Ginters Bušs

August 17, 2009

Abstract

This paper contributes to the literature by comparing predictive accuracy of one-period real-time simple seasonal ARIMA forecasts of Latvia's Gross Domestic Product (GDP) as well as by comparing a direct forecast of Latvia's GDP versus three kinds of indirect forecasts. Four main results are as follows. Direct forecast of Latvia's Gross Domestic Product (GDP) seems to yield better precision than an indirect one. $AR(1)$ model tends to give more precise forecasts than the benchmark moving-average models. An extra regular differencing appears to help better forecast Latvia's GDP in an economic downturn. Finally, only $AR(1)$ gives forecasts with better precision compared to a naïve Random Walk model.

1 Introduction

Most of the macroeconomic forecast literature concentrate on forecasting seasonally adjusted time series (see, for example, Wang (2008), Kuzin *et al.* (2009a, 2009b), Schumacher (2009), Eickmeier and Ng (2009), Boivin and Ng (2006), Diebold and Mariano (1995), Caggiano *et al.* (2009), Masten *et al.* (2009), Barhoumi *et al.* (2009), Forni *et al.* (2003), Dreger and Schumacher (2002), Stock and Watson (1998, 2002, 2003, 2004), and, considering Latvia, Ajevskis and Davidsons (2008) and Benkovskis (2008), among others).

Sometimes, however, it is necessary to forecast seasonally unadjusted series. This paper contributes to the literature by comparing predictive accuracy of one-period real-time simple seasonal ARIMA forecasts of Latvia's Gross Domestic Product (GDP) as well as by comparing a direct forecast of Latvia's GDP versus three kinds of indirect forecasts by predicting its components. The main conclusions in this paper are that a direct forecast of Latvia's GDP outperforms all indirect ones, that an AR(1) process gives more precise forecasts than the competing benchmark models involv-

ing moving-average term, and that only AR(1) model is superior to a naïve Random Walk process.

The paper is organized as follows. Section 2 is meant to be a self-contained description of ARIMA methodology; Section 3 describes the data; Section 4 presents simulation results, and Section 5 concludes.

2 Methodology

2.1 Stochastic processes and stationarity

The purpose of this section is to be sufficiently self-contained. This section up to 2.8 closely follows Kaiser and Maravall (2001) and 2.8 follows Hamilton (1994).

The starting point is the concept of a *stochastic process*. For our purposes, a stochastic process is a real-valued random variable z_t, that follows a distribution $f_t(z_t)$, where t denotes an integer that indexes the period. The T-dimensional variable $(z_{t_1}, z_{t_2}, \ldots, z_{t_T})$ will have a joint distribution that depends on $(t_1, t_2, \ldots, t_T)$. A *time series* $[z_{t_1}, z_{t_2}, \ldots, z_{t_T}]$ will denote a particular realization of the stochastic process. Thus, for each distribution f_t there is only one observation available. Not much can be learned from this, and more structure and more assumptions need to be added. To simplify notation, we shall consider the joint distribution of $(z_1, z_2, \ldots, z_t)$, for which a time series is available for $t \leq T$.

From an applied perspective, the two most important added assumption are

Assumption A : The process is stationary;

Assumption B : The joint distribution of $(z_1, z_2, \ldots, z_t)$ is a multivariate normal distribution.

Assumption A implies the following basic condition. For any value of t,

$$f(z_1, z_2, \ldots, z_t) = f(z_{1+k}, z_{2+k}, \ldots, z_{t+k}) \tag{1}$$

where k is an integer; that is, the joint distribution remains unchanged if all time periods are moved a constant number of periods. In particular, letting $t = 1$, for the marginal distribution it has to be that

$$f_t(z_t) = f(z_t) \tag{2}$$

for every t, and hence the marginal distribution remains constant. This implies

$$Ez_t = \mu_t; \qquad Vz_t = V_z \tag{3}$$

where E and V denote the expectation and the variance operators, respectively, and μ_z and V_z are constants that do not depend on t.

In practice, thus, stationarity implies a constant mean and bounded deviations from it. It is a very strong requirement and few actual economic series will satisfy it. Its usefulness comes from the fact that relatively simple transformations of the non-stationary series will render it stationary. For quarterly economic series, it is usually the case that constant variance can be achieved through the log/level transformation combined with proper outlier correction, and a constant mean can be achieved by differencing.

The log transformation is, roughly speaking, appropriate when the amplitude of the series oscillations increases with the level of the series. As for the outliers, several possible types should be considered, the most popular ones being the additive outlier (i.e., a single spike), the level shift (i.e., a step variable), and the transitory change (i.e., an effect that gradually disappears).

2.2 Differencing

Denote by L the *lag* operator, such that

$$L^j z_t = z_{t-j} \qquad (j = 0, 1, 2, \ldots),$$

and let x_t denote a quarterly observed series. We shall use the operators:

- Regular difference: $\Delta = 1 - L$
- Seasonal difference: $\Delta_4 = 1 - L^4$
- Annual aggregation: $S = 1 + L + L^2 + L^3$

Thus $\Delta x_t = x_t - x_{t-1}$, $\Delta_4 x_t = x_t - x_{t-4}$ and $S x_t = x_t + x_{t-1} + x_{t-2} + x_{t-3}$. It is immediately seen that the 3 operators satisfy the identity

$$\Delta_4 = \Delta S \tag{4}$$

If x_t is a deterministic linear trend, as in $x_t = a + b\,t$, then

$$\Delta x_t = b \tag{5}$$

$$\Delta^2 x_t = 0 \tag{6}$$

where $\Delta^2 x_t = \Delta(\Delta x_t)$. In general, it can easily be seen that Δ^d will reduce a polynomial of degree d to a constant. Obviously, $\Delta_4 x_t$ will also cancel a constant (or reduce the linear trend to a constant); but it will also cancel other deterministic periodic functions, such as one that repeats itself every 4 quarters. To find the set of functions that are cancelled with the transformations $\Delta_4 x_t$. we have to find the solution of the homogeneous difference equation

$$\Delta_4 x_t = (1 - L^4) x_t = x_t - x_{t-4} = 0, \tag{7}$$

with characteristic equation $r^4 - 1 = 0$. The solution is given by

$$r = \sqrt[4]{1}.$$

The four roots are

$$r_1 = 1, \quad r_2 = -1, \quad r_3 = i, \quad r_4 = -i. \tag{8}$$

The first two roots are real and the last two are complex conjugates, with modulus 1 and frequency $\omega = \pi/2$. Complex conjugate roots generate periodic movements of the type

$$r_t = A^t \cos(\omega t + B) \tag{9}$$

where A denotes the amplitude, B denotes the phase (the angle at $t = 0$) and ω - the frequency (the number of full circles that are completed in one unit of time). The period of function (9), to be denoted τ, is the number of units of time it takes for a full circle to be completed, and is related to the frequency ω by the expression

$$\tau = \frac{2\pi}{\omega}. \tag{10}$$

From (8), the general solution of $\Delta_4 x_t = 0$ can be expressed as (see for example, Goldberger, 1967)

$$x_t = c_0 + c_1 \cos(\frac{\pi}{2}t + d_1) + c_2(-1)^t,$$

where c_0, c_1, c_2, and d_1 are constants to be determined from the starting conditions. Realizing that $\cos \pi = -1$, the previous expression can also be written as

$$x_t = c_0 + \sum_{j=1}^{2} c_j \cos(j\frac{\pi}{2}t + d_j), \tag{11}$$

with $d_2 = 0$. Considering (10), the first term in the sum of (11) will be associated with a period of $\tau = 4$ quarters and will represent thus a seasonal component with a once-a-year frequency; the second term has a period of $\tau = 2$ quarters, and hence will represent a seasonal component with a twice-a-year frequency. Noticing that the characteristic equation can be rewritten as $(L^{-1})^4 - 1 = 0$, (8) implies the factorization

$$\Delta_4 = (1 - L)(1 + L)(1 + L^2).$$

The factor $(1-L)$ is associated with the constant and the zero frequency, the factor $(1 + L)$ with the twice-a-year seasonality with frequency $\omega = \pi$, and the factor $(1 + L^2)$ with the once-a-year seasonality with frequency $\omega = \frac{\pi}{2}$. The product of these last two factors yields the annual aggregation operator

S, in agreement with expression (4). Hence, the transformation Sx_t will remove seasonal nonstationarity in x_t.

For the most-often-found case in which stationarity is achieved through the differencing $\Delta\Delta_4$, the factorization

$$\Delta\Delta_4 = \Delta^2 S$$

directly shows that the solution to $\Delta\Delta_4 = 0$ will be of the type:

$$x_t = a + b\,t + \sum_{j=1}^{2} c_j \left[\cos(j\frac{\pi}{2}t) + d_j\right], \tag{12}$$

with $d_2 = 0$. Thus the differencing will remove the same cosine (seasonal) functions as before, plus the local linear trend $(a + bt)$. For the case $\Delta^2\Delta_4$, the factorization $\Delta^3 S$ shows that the canceled trend will now be a second order polynomial in t, the rest remaining unchanged.

A final and important remark:

- Let D denote, in general, the complete differencing applied to the series x_t so as to achieve stationarity. When specifying the ARIMA model for x_t, we shall not be stating that $Dx_t = 0$ but that

  $$Dx_t = z_t,$$

  where z_t is a zero-mean, stationary stochastic process with relatively small variance. Thus every period the solution of $Dx_t = 0$ will be perturbed by the stochastic input z_t (see Box and Jenkins, 1970, Appendix A.4.1). In terms of expression (12), what this perturbation implies is that the a, b, c and d coefficients will not be constant but will instead depend on time. This gradual evolution of the coefficients provides the model with an adaptive behavior that will be associated with the 'moving' features of the trend and seasonal components.

2.3 Linear stationary process, Wold representation, and autocorrelation function

Following the previous notation, if x_t denotes the observed variable and $z_t = Dx_t$ its stationary transformation, under assumptions A and B, the variable $(z_1, z_2, \ldots, z_T)$ will have a proper multivariate normal distribution. One important property of this is that the expectation of some (unobserved) variable linearly related to z_t, conditional on $(z_1, z_2, \ldots, z_T)$, will be a linear function of $z_1, z_2, \ldots, z_T$. Thus conditional expectations will directly provide linear filters. An additional important property is that, because the first two moments fully characterize the distribution, stationarity in mean and variance will imply stationarity of the process. In particular, stationarity

will be implied by the constant mean and variance condition (3), plus the condition that

$$Cov(z_t, z_{t-k}) = \gamma_k,$$

for $k = 0, \pm 1, \pm 2, \ldots$. Hence the covariance between z_t and z_{t-k} should depend on their relative distance k, not on the value of t. Therefore,

$$(z_1, z_2, \ldots, z_T) \sim N(\mu, \Sigma),$$

where μ is a vector of a constant means, and Σ is the variance-covariance matrix

$$\Sigma = \begin{bmatrix} V_z & \gamma_1 & \gamma_2 & \cdots & \gamma_{T-1} \\ & V_z & \gamma_1 & \cdots & \gamma_{T-2} \\ & & \cdots & \cdots & \cdots \\ & & & V_z & \gamma_1 \\ & & & & V_z \end{bmatrix}, \qquad (V_z = \gamma_0),$$

a positive definite symmetric matrix. Let F denote the forward operator, $F = L^{-1}$, such that

$$F^j z_t = z_{t+j}, \quad (j = 0, 1, 2, \ldots),$$

a more parsimonious representation of the 2nd-order moments of the stationary process z_t is given by the Autocovariance Generating Function (AGF)

$$\gamma(L, F) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (L^j + F^j). \tag{13}$$

To transform this function into a scale-free function, we divide by the variance γ_0, and obtain the Autocorrelation Generating Function (ACF),

$$\rho(L, F) = \rho_0 + \sum_{j=1}^{\infty} \rho_j (L^j + F^j), \tag{14}$$

where $\rho_j = \gamma_j / \gamma_0$. If the following conditions on the AGF:

1. $\rho_0 = 1$;
2. $\rho_j = \rho_{-j}$;
3. $|\rho_j| < 1$ for $j \neq 0$;
4. $\rho_j \to 0$ as $j \to \infty$;
5. $\sum_{j=0}^{\infty} |\rho_k| < \infty$,

are satisfied, then a zero-mean, finite-variance, normally distributed process is stationary. Further, under the normality assumption, a complete realization of the stochastic process will be fully characterized by μ_z, V_z and $\rho(L, F)$.

When $\rho_j = 0$ for all $j \neq 0$, the process will be denoted a White Noise process. Therefore, a white noise process is a sequence of normally identically independently distributed random variables.

The first statistics that we shall compute for a time series $[z_1, z_2, \ldots, z_T]$ will be estimates of the autocovariances and autocorrelations using the standard sample estimates

$$\bar{z} = T^{-1} \sum_{t=1}^{T} z_t; \qquad \hat{\gamma}_k = T^{-1} \sum_{t=k+1}^{T} (z_t - \bar{z})(z_{t-k} - \bar{z}); \qquad \hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0.$$

To start the modeling procedure, a general result on linear time series processes will provide us with an analytical representation of the process that will prove very useful. This is the so-called Wold representation. We present it next.

Let z_t denote a linear stationary stochastic process with no deterministic component, then z_t can be expressed as the one-sided moving average

$$\begin{aligned} z_t &= a_t + \ {}_1a_{t-1} + \ {}_2a_{t-2} + \ldots = \\ &= \sum_{j=0}^{\infty} \ {}_ja_{t-j} = \Psi(L)a_t, \qquad \text{where} \end{aligned}$$

$$\Psi(L) = \sum_{j=0}^{\infty} \ {}_jL^j, \qquad (\ {}_0 = 1), \tag{15}$$

where a_t is a white noise process with zero mean and constant variance V_a, and $\Psi(L)$ is such that

1. $\ {}_j \to 0$ as $j \to \infty$;
2. $\sum_{j=0}^{\infty} |\ {}_j| < \infty$;

the last condition reflecting a sufficient condition for convergence of the polynomial $\Psi(L)$. Given the $\ {}_j$-coefficients, a_t represents the one-period ahead forecast error of z_t, that is

$$a_t = z_t - \hat{z}_{t|t-1},$$

where $\hat{z}_{t|t-1}$ is the forecast of z_t made at period $t-1$. Since a_t represents what is new in z_t, it will be referred to as the *innovation* of the process. The representation of z_t in terms of its innovations, given by (15), is unique, and is usually referred to as the Wold representation.

A useful result is the following: If $\gamma(L, F)$ represents the AGF of the process z_t, then

$$\gamma(L, F) = \Psi(L)\Psi(F)V_a. \tag{16}$$

In particular, for the variance,

$$V_z = (1 + \ {}^{2}_{1} + \ {}^{2}_{2} + \ldots)V_a. \tag{17}$$

2.4 The spectrum

The spectrum is the basic tool in the so-called 'Frequency Domain Approach' to time series analysis. It represents an alternative way to look at and interpret the information contained in the second-order moments of the series. Here we provide a few basic concepts needed in later subsections.

Consider, first, a time series given by $z_1, z_2, \ldots, z_T$. To simplify the discussion, assume the process has a zero mean and that T is even, so that one can write $T = 2q$. In the same way that, as is well known, the T values of z_t can be exactly duplicated by a polynomial of order $(T-1)$, they can also be exactly reproduced as the sum of $T/2$ cosine functions of the type (9); this result, in fact, provides the basis for Fourier analysis.

We start by defining the Fundamental Frequency $\omega = 2\pi/T$ (i.e., the frequency of one full circle completed in T periods) and its multiples (or *harmonics*) $\omega_j = (2\pi/T)j$, $j = 1, 2, \ldots, q$. Then, express (9) as

$$r_{jt} = a_j \, cos\, \omega_j t + b_j \, sin\, \omega_j t, \tag{18}$$

and hence,

$$z_t = \sum_{j=1}^{q} r_{jt}. \tag{19}$$

It is straightforward to check that a_j and b_j are related to the amplitude A_j by $A_j^2 = a_j^2 + b_j^2$. From (18) and (19), by plugging in the values of z_t, ω_j, and t, a linear system of T equations is obtained in the unknowns a_j's and b_j's, $j = 1, 2, \ldots, q$; a total of T unknowns. Therefore, for each frequency ω_j, we obtain a square amplitude A_j^2. As a consequence, we obtain a set of periodic functions with different frequencies and amplitudes. We can group the functions in intervals of frequency by summing the squared amplitudes of the functions that fall in the same interval. In this way one obtains a histogram of frequencies that shows the contribution of each interval of frequency to the series variation. In the same way that a density function is the model counterpart of the usual histogram, the spectrum is the model counterpart of the frequency histogram, properly standardized.

We can now let the interval $\Delta\omega_j$ go to zero, and the frequency histogram becomes a continuous function, which is denoted as the *sample spectrum*. The area over the differential $d\omega$ represents the contribution of the frequencies in $d\omega$ to the variation of the time series. An important result links

the sample spectrum with the SACF. If $H(\omega)$ denotes the sample spectrum, then it is proportional to

$$H(\omega) \propto \quad \hat{\gamma}_0 + 2\sum_{t=1}^{T-1} \hat{\gamma}_j \, cos\, \omega t \Bigg), \tag{20}$$

where $\hat{\gamma}_j$ denotes the lag-j autocovariance estimator.

The model equivalent of (20) provides precisely the definition of the power spectrum (see below). Consider the AGF of the stationary process z_t, given by

$$\gamma(L, F) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (L^j + F^j), \tag{21}$$

where L is a complex number of unit modulus, which can be expressed as $e^{i\omega}$. Replacing L and F by their complex representation, (21) becomes the function

$$g(\omega) = \gamma_0 + \sum_{j=1}^{\infty} \gamma_j (e^{-i\omega j} + e^{i\omega j}),$$

or, using the identity $\left[e^{-i\omega j} + e^{i\omega j} = 2\, cos(j\omega)\right]$, and dividing by 2π, one obtains

$$g_1(\omega) = \frac{1}{2\pi} \left[\gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j \, cos(j\omega) \right]. \tag{22}$$

The move from (21) to (23) is the so-called Fourier cosine transform of the AGF $\gamma(L, F)$, and is termed the *power spectrum*. Replacing the AGF by the ACF (i.e., dividing by the variance γ_0), we obtain the Spectral Density Function

$$g_1^*(\omega) = \frac{1}{2\pi} \left[1 + 2\sum_{j=1}^{\infty} \rho_j \, cos(j\omega) \right]. \tag{23}$$

It is easily seen that $g_1(\omega)$ and $g_1^*(\omega)$ are periodic functions, and hence the range of frequencies can be restricted to $(-\pi, \pi)$ or $(0, 2\pi)$. Moreover, given that the cosine function is symmetric around zero, we only need to consider the range $(0, \pi)$.

From (23), knowing the AGF of a process, the power spectrum is trivially obtained. Alternatively, knowledge of the power spectrum permits us to derive the AGF by means of the inverse Fourier transform, given by

$$\gamma_k = \int_{-\pi}^{\pi} g(\omega) \, cos(\omega k) \, d\omega.$$

Thus, for $k = 0$,

$$\gamma_0 = \int_{-\pi}^{\pi} g(\omega) \, d\omega, \tag{24}$$

which shows that the integral of the power spectrum is the variance of the process.

As an example, consider a process z_t, the output of the 2^{nd}-order homogeneous difference-equation model

$$z_t + .81 z_{t-2} = 0 \tag{25}$$

The characteristic equation, $r^2 + .81 = 0$ yields the pair of the complex conjugate numbers $r = \pm .9i$, situated in the imaginary axis, they will be associated thus with the frequency $\omega = \pi/2$. The process, therefore, follows the deterministic function

$$z_t = .9\, cos\big(\frac{\pi}{2} t + \beta\big), \tag{26}$$

where we can set $\beta = -\pi/2$. The function (26) does not depend on ω and the movements of z_t are all associated with single frequency $\omega = \pi/2$. To transform the previous model into a stochastic process, we perturb the equilibrium (25) every period with a white noise $(0, 1)$ variable a_t, so that it is replaced by the stochastic model

$$z_t + .81 z_{t-2} = a_t, \quad \text{or} \quad (1 + .81 L^2) z_t = a_t. \tag{27}$$

From (27), the Wold representation (15) is immediately obtained through

$$\gamma(L, F) = \frac{V_a}{(1 + .81L^2)(1 + .81F^2)} = \\ = \frac{V_a}{1.656 + .81(L^2 + F^2)}.$$

Replacing $(L^2 + F^2)$ by $2\, cos\, 2\omega$, the spectrum is found to be equal to

$$g(\omega) = \frac{V_a}{1.656 + 1.62\, cos\, 2\omega}; \qquad 0 \le \omega \le \pi.$$

In summary, if a series contains an important component for a certain frequency ω_0, its spectrum should reveal a peak around that frequency.

2.5 ARIMA models

Back to the Wold representation (15) of a stationary process, $z_t = \Psi(L) a_t$. This representation is of no help from the point of view of fitting a model because, in general, the polynomial $\Psi(L)$ will contain an infinite number of parameters. Therefore, we use a rational approximation of the type

$$\Psi(L) \doteq \frac{\theta(L)}{\phi(L)},$$

where $\theta(L)$ and $\phi(L)$ are finite polynomials in L of order q and p, respectively. Then we can write

$$z_t = \frac{\theta(L)}{\phi(L)} a_t, \text{ or}$$

$$\phi(L) z_t = \theta(L) a_t. \tag{28}$$

The model

$$(1 + \phi_1 L + \ldots + \phi_p L^p) z_t = (1 + \theta_1 L + \ldots + \theta_q L^q) a_t \tag{29}$$

is the Autoregressive Moving-Average process of orders p and q; in brief, the ARMA(p, q) model. For further reference, the *inverse model* of (28) is the one that results from interchanging the AR and MA polynomials. Thus

$$\theta(L) y_t = \phi(L) b_t,$$

with b_t white noise, is an inverse model of (28). Equation (29) can be seen as a non-homogeneous difference equation with forcing function $\theta(L)a_t$, an MA(q) process. Therefore, if both sides of (29) are multiplied by z_{t-k}, with $k > q$, and expectations are taken, the right hand side of the equation vanishes, and the left hand side becomes:

$$\gamma_k + \phi_1 \gamma_{k-1} + \ldots + \phi_p \gamma_{k-p} = 0, \tag{30}$$

or

$$\phi(L) \gamma_k = 0, \tag{31}$$

where L operates on the subindex k. The eventual autocorrelation function (that is, γ_k as a function of k, for $k > q$) is the solution of the homogeneous difference equation (30), with characteristic equation

$$r_p + \phi_1 r^{p-1} + \ldots + \phi_p = 0. \tag{32}$$

If $r_1, \ldots, r_p$ are the roots of (32), the solution of (30) can be written as

$$\gamma_k = \sum_{i=1}^{p} r_i^k,$$

and will converge to zero as $k \to \infty$ when $|r_i| < 1$, $i = 1, \ldots, p$. Comparison of (32) and (30) shows that $r_1, \ldots, r_p$ are the inverses of the roots $L_1, \ldots, L_p$ of the polynomial

$$\phi(L) = 0$$

that is, $r_i = L_i^{-1}$. Convergence of γ_k implies, thus, that the roots (in L) of the polynomial $\phi(L)$ are all larger than 1 in modulus. This condition can also be stated as follows: the roots of the polynomial $\phi(L)$ have to lie

outside the unit circle. When this happens, it is said that the polynomial $\phi(L)$ is stable. From the identity

$$\phi(L)^{-1} = \frac{1}{(1 - r_1 L) \dots (1 - r_p L)}$$

it is seen that stability of $\phi(L)$ implies, in turn, convergence of its inverse $\phi(L)^{-1}$.

From (21), considering that $\Psi(L) = \theta(L)/\phi(L)$, the AGF of z_t is given by

$$\gamma(L, F) = \frac{\theta(L)}{\phi(L)} \frac{\theta(F)}{\phi(F)} V_a \tag{33}$$

and it is straightforward to see that the stability of $\phi(L)$ will imply that the stationarity conditions of Section 2.3 are satisfied. The AGF is symmetric and convergent, and the eventual autocorrelation function is the solution of a difference equation, and hence, in general, a mixture of damped polynomials in time and periodic functions. The Fourier transform of (33) yields the spectrum of z_t, equal to

$$g_z(\omega) = V_a \frac{\theta(e^{-i\omega})\theta(e^{i\omega})}{\phi(e^{-i\omega})\phi(e^{i\omega})}, \tag{34}$$

and the integral of $g_z(\omega)$ over $0 \leq \omega \leq 2\pi$ is equal to $2\pi\, Var(z_t)$.

A useful result is the following. If two stationary stochastic processes are related through

$$y_t = C(L)x_t,$$

then the AGF of y_t, $\gamma_y(L, F)$ is equal to

$$\gamma_y(L, F) = C(L)C(F)\gamma_x(L, F),$$

where $\gamma_x(L, F)$ is the AGF of x_t. Finally, a function that will prove helpful is the Crosscovariance Generating Function (CGF) between two series, x_t and y_t, with Wold representation

$$x_t = \alpha(L)a_t$$

$$y_t = \beta(L)a_t.$$

Letting $\gamma_j = E(x_t y_{t-j})$ denote the lag-j crosscovariance between x_t and y_t, $j = 0, \pm 1, \pm 2, \dots$, the CGF is given by

$$CGF(L, F) = \sum_{-\infty}^{\infty} \gamma_j L^j = \alpha(L)\beta(F)\sigma_a^2.$$

If, in equation (29), the subindex t is replaced by $t+k$ (k a positive integer), and expectations are taken at time t, the forecast of z_{t+k} made at time t,

namely $\hat{z}_{t+k|t}$ is denoted the *forecast function*. Given that $E_t a_{t+k} = 0$ for $k > 0$, it is found that, for $k > q$, the forecast function satisfies the equation

$$\hat{z}_{t+k|t} + \phi_1 \hat{z}_{t+k-1|t} + \ldots + \phi_p \hat{z}_{t+k-p|t} = 0,$$

where $\hat{z}_{t+j|t} = z_{t+j}$ when $j \leq 0$. Therefore, the eventual forecast function is the solution of

$$\phi(L)\hat{z}_{t+k|t} = 0, \tag{35}$$

with L operating on k. Comparing (31) and (35), the link between autocorrelation for lag k (and longer) and k-period-ahead forecast becomes apparent, the forecast being simply an extrapolation of correlation: what we can forecast is the correlation we have detected. For a zero-mean stationary process the forecast function will converge to zero, following, in general, a mixture of damped exponentials and cosine functions.

In summary, stationarity of an ARMA model, which requires the roots of the autoregressive polynomial $\phi(L)$ to be larger than 1 in modulus, implies the following model properties: a) its AGF converges; b) its forecast function converges; and c) the polynomial $\phi(L)^{-1}$ converges, so that z_t accepts the convergent (infinite) MA representation

$$z_t = \phi(L)^{-1}\theta(L)a_t = \Phi(L)a_t, \tag{36}$$

which is precisely the Wold representation. For example, for the AR(1) model

$$z_t + \phi z_{t-1} = a_t,$$

the root of $1 + \phi L = 0$ is $L_1 = -1/\phi$. Thus, stationarity of z_t implies that $|L_1| = |\frac{1}{\phi}| > 1$, or $|\phi| < 1$.

If z_t is the differenced series, for which stationarity can be assumed, that is

$$z_t = Dx_t, \qquad D = \Delta^d, \quad d = 0, 1, 2, \ldots,$$

then the original nonstationary series x_t follows the Autoregressive Integrated Moving-Average process of orders p, d, and q, or ARIMA(p, d, q) model, given by

$$\phi(L)Dx_t = \theta(L)a_t; \tag{37}$$

p and q refers to the orders of the AR and MA polynomials, respectively, and d refers to the number of regular differences (i.e., the number of unit roots at the zero frequency). In the following, we might encounter the following abbreviations:

AR(p): autoregressive process of order p ;

MA(q): moving-average process of order q ;

ARI(p, d): autoregressive process of order p applied to the d^{th} difference of the series;

IMA(d, q): moving-average process of order q applied to the d^{th} difference of the series;

Further, a series will be denoted I(d) when it requires d regular differences in order to become stationary.

As in the stationary case, taking conditional expectations at time t in both sides of equation (37) with t repaced by $t + k$, where k is a positive integer, it is obtained that

$$\phi(L)D\hat{x}_{t+k|t} = \theta(L)\hat{a}_{t+k|t},$$

where $\hat{x}_{t+j|t} = E(x_{t+j}|x_t, x_{t-1}, \ldots)$ is the forecast of x_{t+j} obtained at time t when $j > 0$, and is the observation x_{t+j} when $j \leq 0$; further, $\hat{a}_{t+j|t} = E(a_{t+j}|x_t, x_{t-1}, \ldots)$ is equal to 0 when $j > 0$, and is equal to a_{t+j} when $j \leq 0$. As a consequence, the eventual forecast function ($\hat{x}_{t+k|t}$ as a function of k, for $k > q$) will be the solution of the homogeneous difference equation

$$\phi(L)D\hat{x}_{t+k|t} = 0,$$

with L operating on k. The roots of L all have unit modulus; if $L = \Delta^d$, then the eventual forecast function will include a deterministic polynomial in t of the type $(a + b\, t^{d-1})$. If L includes also seasonal differencing Δ_4, then the eventual forecast function will contain also the non-convergent deterministic cosine-type function (11), associated with the once and twice-a-year seasonal frequencies, $\omega = \pi/2$ and $\omega = \pi$.

As an example, the forecast function of the model

$$(1 - .7L)\Delta\Delta_4 x_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t,$$

will consist of five starting values $\hat{x}_{t+j|j}$, $j = 1, \ldots, 5$, implied by the MA part with $q = 5$, after which the function will be the solution of the homogeneous equation associated with the AR part. Factorizing the AR polynomial as

$$(1 - .7L)(1 - L)^2(1 + L)(1 + L^2),$$

the roots of the characteristic equation are given by $r_1 = .7$, $r_2 = r_3 = 1$, $r_4 = -1$, $r_5 = i$, $r_6 = -i$. From Section 2.2, the eventual forecast function can be expressed as

$$\hat{x}_{t+k|t} = c_1^{(t)}(.7)^k + c_2^{(t)} + c_3^{(t)}k + c_4^{(t)}(-1)^k + c_5^{(t)}cos\left(\frac{\pi}{2}k + c_6^{(t)}\right),$$

where the last two terms reflect the seasonal harmonics (the root $r_4 = -1$ can also be written as $c_4^{(t)}cos\pi k$). The constants $c_1, \ldots, c_6$ are determined from the starting conditions of the forecast function, and hence will depend on t, the origin of the forecast. This feature gives the ARIMA model its adaptive (or 'moving') properties. Notice that, in the nonstationary case,

the forecast function (with fixed origin t and increasing horizon k) will not converge.

Concerning the MA polynomial $\theta(L)$, a similar condition of stability will be imposed, namely, the roots $L_1, \ldots, L_q$ of the equation $\theta(L) = 0$ have to be larger than 1 in modulus. This condition is referred to as the *invertibility condition* for the process and, unless otherwise specified, we shall assume that the model for the observed series z_t is invertible. This assumption implies that $\theta(L)^{-1}$ converges, so that the model (28) can be inverted and expressed as

$$a_t = \theta(L)^{-1}\phi(L)z_t = \Pi(L)z_t, \tag{38}$$

which shows that the series accepts a convergent (infinite) AR expression, and hence can be approximated by a finite AR. Expression (38) also shows that, when the process is invertible, the innovations can be recovered from the z_t series.

Some frequency domain implications of nonstationarity and noninvertibility are worth pointing out. Assume that the MA polynomial $\theta(L)$ has a unit root $|L_1| = 1$ - perhaps a complex conjugate pair - associated with the frequency ω_1. Then, $\theta(e^{-i\omega_1}) = 0$, and the spectrum of z_t, given by (34), will have a zero for the frequency ω_1. Analogously, if $|L_1| = 1$ is a root of the AR polynomial $\phi(L)$, with associated frequency ω_1, then, $\phi(e^{-i\omega_1}) = 0$ and $g(\omega_1) \to \infty$.

It follows that

- a unit MA root causes a zero in the spectrum;
- a unit AR root causes a point of ∞ in the spectrum;
- an invertible model will have strictly positive spectrum, $g(\omega) > 0$;
- a stationary model has a bounded spectrum, $g(\omega) < \infty$.

For quarterly data with seasonality, the differencing L is likely to contain the seasonal difference Δ_4. A popular specification that increases parsimony of the model and permits us to capture seasonal effects is the multiplicative seasonal model

$$\phi(L)\Phi(L^4)\Delta^d\Delta_4^D x_t = \theta(L)\Theta(L^4)a_t, \tag{39}$$

where the regular AR polynomial in L, $\phi(L)$, is as in (29), $\Phi(L^4)$ is the seasonal AR polynomial in L^4, d is the degree of regular differencing, D is the degree of seasonal differencing, $\theta(L)$ is the regular MA polynomial in L, $\Theta(L^4)$ is the seasonal MA polynomial in L^4, and a_t denotes the series white-noise $(0, V_a)$ innovation. The polynomials $\phi(L)$, $\Phi(L^4)$, $\theta(L)$ and $\Theta(L^4)$ are assumed stable, and hence the series

$$z_t = \Delta^d\Delta_4^D x_t$$

follows a stationary and invertible process. If p, P, q, and Q denote the orders of the respective polynomials, model (39) will be referred to as the multiplicative ARIMA$(p,d,q)(P,D,Q)_4$ model. In practice, we can safely restrict the orders to

$$\begin{array}{rl} p,q & \leq 4 \\ P & \leq 1 \\ Q & \leq 2 \\ d & \leq 2 \\ D & \leq 1. \end{array} \tag{40}$$

Two important practical moments are the following:

1. Parsimony (i.e., few parameters) should be a crucial property of ARIMA models used in practice.

2. ARIMA models are a useful tool for relatively short-term analysis. Their flexibility and adaptive behavior contribute to their good short-term forecasting. Tong-term extrapolation of this flexibility may imply, however, ustable long-term inference (see, for example, Maravall, 1999). As a general rule, short-term analysis favors differencing, while long-term one favors more deterministic trends, that implies less differencing.

2.6 Modeling strategy, diagnostics and inference

The so-called Box-Jenkins approach to building ARIMA models consists of the following iterative scheme that contains 4 stages.

2.6.1 Identification

Two features of the series have to be addressed:

- the degree of regular and seasonal differencing;
- the orders of the stationary AR and invertible MA polynomials.

Differencing of the series can employ some of the unit root tests available for possibly seasonal data (see, for example, Hylleberg *et al*,1990). Devised to test deterministic seasonals versus seasonal differencing, these tests are of little use for our purpose. Consider two models:

(a) $x_t = \mu + a_t$,

(b) $\Delta x_t = (1 - .99L)a_t$.

For a quarterly series, and realistic series length, it is impossible that the sample information can distinguish between the two specifications. Consequently, the choice is arbitrary. Besides the variance of a_t, model (a) contains one parameter that needs to be estimated, while model (b) contains none (although, in this case the first observation is lost by differencing). Model (a) offers, thus, no estimation advantage. If short-term forecasting is the main objective, however, model (b) will display some advantage because it allows for more flexibility, given that it could be rewritten as $x_t = \mu_{(t)} + a_t$, where $\mu_{(t)}$ is a very slowly adapting mean.

A similar consideration applies to seasonal variations. The model

(c) $x_t = \mu + \sum_{j=1}^{3} \beta_j d_{jt} + a_t$,

where d_{jt} denotes a quarterly seasonal dummy variable, is in practice indistinguishable from the direct specification

(d) $\Delta_4 x_t = (1 - .95L^4)a_t$.

The deterministic specification has now 4 parameters; the stochastic one has none, but four starting values are lost at the beginning. The latter can also be expressed as

$$x_t = \mu^{(t)} + \sum_{j=1}^{3} \beta_j^{(t)} d_{jt} + a_t,$$

where $\mu^{(t)}$ and $\beta_j^{(t)}$ denote slowly adapting coefficients. Within our short-term perspective, there is no reason thus to maintain the deterministic-stochastic dichotomy, and deterministic features can be seen as extremely stable stochastic ones.

Besides the lack of the power of unit root tests to distinguish between models (a) and (b), or (c) and (d), the process of building ARIMA models typically implies estimation of many specifications. In practice, a more efficient and reliable procedure for determining AR roots is to use estimation results based on the superconsistency of parameter estimates associated with unit roots, having determined *a priori* how close to unity a root has to be in order to be considered a unit root (see Tiao and Tsay, 1983, 1989, and Gómez and Maravall, 2000a).

Once the proper differencing has been established, it remains to determine the orders of the stationary AR and invertible MA polynomials. Here, the basic criterion used to be to try to match the SACF of z_t with the theoretical ACF of a particular ARMA process. In recent years, the efficiency and reliability of automatic identification procedures, based mostly on information criteria, has strongly decreased the importance of the 'tentative identification' stage (see Fischer and Planas, 1999, and Gómez and Maravall, 2000a).

2.6.2 Estimation and diagnostics

When $q \neq 0$, the ARIMA residuals are highly nonlinear functions of the model parameters, and hence numerical maximization of the likelihood, or of some function of the residual sum of squares, can be computationally non-trivial. Within the restrictions in the size of the model given by (40), however, maximization is typically well behaved. A standard estimation procedure would cast the model in a state-space format, and use the Kalman filter to compute the likelihood through the Prediction Error Decomposition. The likelihood is then maximized with some nonlinear procedure. Usually, the V_a parameter, as well as a possible constant mean, are concentrated out of the likelihood (see Section 2.8). When the series is nonstationary, several solutions have been proposed to overcome the problem of defining a proper likelihood. Relevant references are Bell and Hillmer (1991), Brockwell and Davis (1987), De Jong (1991), Gómez and Maravall (1994), Kohn and Ansley (1986), and Morf, Sidhu and Kailath (1974).

Many diagnostics are available for ARIMA models. A crucial one, of course, is the out-of-sample forecast performance. Some tests for in-sample model stability are also of interest. Also, there is a large set of tests based on the model residuals, assumed to be *niid*. This implies testing for normality, autocorrelation, homoskedasticity, etc. Besides the ones proposed by Box and Jenkins (1970), additional references can be Newbold (1983), Gourieroux and Monfort (1990), Harvey (1989), and Hendry (1995).

2.6.3 Inference

If the diagnostics are failed, in the light of the results obtained, the model specification should be changed. When the model passes all diagnostics, we may then proceed to inference. We shall look in particular at an application in forecasting, the topic of this paper.

Let (37) denote, in compact notation, the ARIMA model identified for the series x_t , and denote by $\hat{x}_{t+j|t}$ the forecast of x_{t+j} made at point t (in Box-Jenkins notation, $\hat{x}_{t+j|t} = \hat{x}_t(j)$). Under our assumptions, the optimal forecast of x_{t+j}, in Minimum Mean Square Error (MMSE) sense, is the expectation of x_{t+k} conditional on the observed time series $x_1, \ldots, x_t$ (equal also to the projection of x_{t+k} onto the observed series); that is,

$$\hat{x}_{t+j|t} = E(x_{t+j}|x_1, \ldots, x_t).$$

Recall that, for known parameters,

$$a_t = x_t - \hat{x}_{t|t-1},$$

that is, the innovations of the process are the sequence of one-period-ahead forecast errors.

The forecast function at time t is $\hat{x}_{t+k|t}$ as a function of k (k a positive integer). In Section 2.5 we saw that for an ARIMA(p, d, q) model, the forecast function consists of q starting conditions, after which it is given by the solution of the homogeneous AR difference equation

$$\phi^*(L)\hat{x}_{t+k|t} = 0, \tag{41}$$

where L operates on k, and $\phi^*(L)$ denotes the full AR convolution $\phi^*(L) = \phi(L)D$, and includes thus the unit roots.

A useful way to look at forecasts is directly based on the pure MA representation $\Psi(L)$, even in the nonstationary case of a nonconvergent $\Psi(L)$. Assume the model parameters are known and write

$$x_{t+k} = a_{t+k} + \ {}_1a_{t+k-1} + \ldots + \ {}_{k-1}a_{t+1} + \ {}_ka_t + \ {}_{k+1}a_{t-1} + \ldots. \tag{42}$$

Given that, for $k > 0$, $E_t a_{t+k} = 0$ and $E_t a_{t-k} = a_{t-k}$, taking conditional expectations in (42) yields

$$\hat{x}_{t+k|t} = E_t x_{t+k} = \sum_{j=0}^{\infty} \ {}_{k+j}a_{t-j}; \tag{43}$$

so that the forecast is a linear combination of past and present innovations. Subtracting (43) from (42), the k-periods-ahead forecast error is given by the model

$$\begin{aligned} e_{t+k|t} &= x_{t+k} - \hat{x}_{t+k|t} \\ &= a_{t+k} + \ {}_1a_{t+k-1} + \ldots + \ {}_{k-1}a_{t+1}, \end{aligned} \tag{44}$$

an MA$(k-1)$ process of 'future' innovations. From expression (44), the joint, marginal, and conditional distributions of forecast errors can be easily derived, and in particular the standard error of the k-period ahead forecast, equal to

$$SE(k) = (1 + \ {}_1^2 + \ldots + \ {}_{k-1}^2)^{1/2}\sigma_a. \tag{45}$$

Unless the series is relatively short, this standard error, estimated by using ML estimators of the parameters, will provide a good approximation.

2.6.4 A particular class of models

Box and Jenkins (1970) dedicate a considerable amount of attention to a particular multiplicative model that, for quarterly series, takes the form

$$\Delta\Delta_4 x_t = (1 + \theta_1 L)(1 + \theta_4 L^4)a_t \tag{46}$$

(a regular IMA$(1, 1)$ structure multiplied by a seasonal IMA$(1, 1)$ structure). Given that they identified the model for a series of airline passengers, it has

become know as the 'Airline model'. Often, the model is obtained for the logs, in which case a rough first reading shows that the rate-of-growth of the annual difference is a stationary process.

The model is highly parsimonious, and the 3 parameters can be given a structural interpretation. As seen in Section 2.5, when $\theta_1 \to -1$, the trend behavior generated by the model becomes more and more stable and, when $\theta_4 \to -1$, the same thing happens to the seasonal component. Estimation of MA roots close to the noninvertibility boundary poses no serious problem, and fixing *a priori* the maximum value of the modulus of an MA root to, for example, .99 produces perfectly behaved invertible models.

If estimation of (46) yields, for example, $\hat{\theta}_4 = -.99$, two (mutually exclusive) things can explain the result:

1. seasonality is practically deterministic;

2. there is no seasonality, and the model is overdifferenced.

Determining which of the two is the correct explanation is rather simple by testing for the significance of seasonal dummy variables. When the model has no seasonality, the seasonal filter $\Delta_4 z_t = (1 - .99L^4)b_t$ would have hardly any effect on the input series. A similar reasoning holds for θ_1 and the possible presence of a deterministic trend. Further, a purely white-noise series filtered with model (46) with $\theta_1 = \theta_4 = -.99$ would, very approximately, reproduce the series. Thus, the Airline model also encompasses simpler structures with no trend or no seasonality. Adding the empirical fact that it provides reasonably good fits to many actual macroeconomic series (see, for example, Fischer and Planas, 1999, or Maravall, 2000) it is an appropriate model for illustration, for benchmark comparison, and for pre-testing.

2.7 Preadjustment

We have introduced the ARIMA model as a practical way of dealing with moving features of series. Still, before considering a series appropriate for ARIMA modelling, several prior corrections or adjustments may be needed. We shall classify them in 3 groups.

1. Outliers

   The series may be subject to abrupt changes, that cannot be explained by the underlying normality of the ARIMA model. Three main types of outlier effects are often distinguished: a) *additive outliers*, which affects an isolated observation, b) *level shift*, which implies a step change in the mean level of the series, and c) *transitory change*, similar to an additive outlier whose effect damps out over a few periods. Chen and Liu (1993) suggested an approach to automatic outlier detection that

has been implemented by TRAMO/SEATS program (see Gómez and Maravall, 2000a).

2. Calendar effect

   By this term we refer to the effect of calendar dates, such as the number of working days in a period, the location of Easter effect, or holidays. These effects are typically incorporated into the model through regression variables (see, for example, Hillmer, Bell and Tiao, 1983, and Harvey, 1989).

3. Invervention variables

   Often special, unusual events affect the evolution of the series and cannot be accounted for by the ARIMA model. There is thus a need to 'intervene' the series in order to correct for the effect of special events. Examples can be strikes, devaluations, change of the base index or of the way a series is constructed, natural disasters, political events, important tax changes, or new regulation, to mention a few. These special effects are entered in the model as a regression variables.

The full model for the observed series can thus be written as

$$y_t = w_t'\beta + C_t'\eta + \sum_{j=1}^{k} \alpha_j \lambda_j(L) I_t(t_j) + x_t, \tag{47}$$

where $\beta = (\beta_1, \ldots, \beta_n)'$, is a vector of regression coefficients, $w_t' = (w_{1t}, \ldots, w_{nt})$ denotes n regression or intervention variables, C_t' denotes the matrix with columns of the calendar effects' variables (trading day, Easter effect, Leap year effect, holidays), and η the vector of associated coefficients, $I_t(t_j)$ is an indicator variable for the possible presence of an outlier at period t_j, $\lambda_j(L)$ captures the transmission of the j-th outlier effect (for additive outliers, $\lambda_j(L) = 1$, for level shifts, $\lambda_j(L) = 1/\Delta$, for transitory changes, $\lambda_j(L) = 1/(1 - \delta L)$, with $0 < \delta < 1$,) and α_j denotes the coefficient of the outlier in the multiple regression model with k outliers. Finally, x_t follows a general (possibly multiplicative) ARIMA model (39). As mentioned earlier, there are several procedures for estimation of models of this type, and easily available programs that enforce the procedures (for example, TRAMO; see Gómez and Maravall, 1996). Noticing that intervention variables, outliers and calendar effects are regression variables, the full model can be expressed as a regression-ARIMA model which, in our case, is estimated by the Kalman filter that we discuss next.

2.8 Kalman filter

This section follows Hamilton (1994) and introduces some very useful tools named for the contribution of R.E. Kalman (1960, 1963). The idea is to ex-

press a dynamic system in a particular form called the *state-space representation*. The Kalman filer is an algorithm for sequentially updating a linear projection for the system. Among other benefits, this algorithm provides a way to calculate exact finite-sample forecasts and the exact likelihood function for Gaussian ARMA processes, to factor matrix autocovariance-generating functions or spectral densities, and to estimate vector autoregressions with coefficients that change over time.

Subsection 2.8.1 describes how a dynamic system can be written in a form that can be analyzed using the Kalman filter. The filter itself is derived in Subsection 2.8.2, and its use in forecasting is described in Subsection 2.8.3. Subsection 2.8.4 explains how to estimate the population parameters by maximum likelihood. Subsection 2.8.5 develops a smoothing algorithm, which is a way to use all the information in the sample to form the best inference about the unobserved state of the process at any historical date. Finally, Subsection 2.8.6 describes standard errors for smoothed inferences and forecasts.

2.8.1 The State-Space Representation of a Dynamic System

Maintained Assumptions Let y_t denote an $(n \times 1)$ vector of variables observed at date t. A rich class of dynamic models for y_t can be described in terms of a possibly unobserved $(r \times 1)$ vector ξ_t known as the *state vector*. The *state-space representation* of the dynamics of y is given by the following system of equations:

$$\xi_{t+1} = F\xi_t + v_{t+1} \tag{48}$$

$$y_t = A'x_t + H'\xi_t + w_t, \tag{49}$$

where F, A', and H' are matrices of parameters of dimension $(r \times r)$, $(n \times k)$, and $(n \times r)$, respectively, and x_t is a $(k \times 1)$ vector of exogenous or predetermined variables. Equation (48) is known as the *state equation*, and (49) is known as the *observation equation*. The $(r \times 1)$ vector v_t and the $(n \times 1)$ vector w_t are vector white noise:

$$E(v_t v_\tau') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{50}$$

$$E(w_t w_\tau') = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise,} \end{cases} \tag{51}$$

where Q and R are $(r \times r)$ and $(n \times n)$ matrices, respectively. The disturbances v_t and w_t are assumed to be uncorrelated at all lags:

$$E(v_t, w_\tau') = 0 \qquad \text{for all } t \text{ and } \tau. \tag{52}$$

The statement that x_t is predetermined or exogenous means that x_t provides no information about ξ_{t+s} or w_{t+s} for $s = 0, 1, 2, \ldots$ beyond that contained

in $y_{t-1}, y_{y-2}, \ldots, y_1$. Thus, for example, x_t could include lagged values of y or variables that are uncorrelated with ξ_τ and w_τ for all τ.

The system of (48) through (52) is typically used to describe a finite series of observations $\{y_1, y_2, \ldots, y_T\}$ for which assumptions about the initial value of the state vector ξ_1 are needed. We assume that ξ_1 is uncorrelated with any realizations of v_t and w_t:

$$E(v_t \xi_1') = 0 \qquad \text{for } t = 1, 2, \ldots, T \tag{53}$$

$$E(w_t \xi_1') = 0 \qquad \text{for } t = 1, 2, \ldots, T \tag{54}$$

The state equation (48) implies that ξ_t can be written as a linear function of $(\xi_1, v_2, v_3, \ldots, v_t)$:

$$\xi_t = v_t + F v_{t-1} + F^2 v_{t-2} + \ldots + F^{t-2} v_2 + F^{t-1} \xi_1 \qquad \text{for } t = 2, 3, \ldots, T. \tag{55}$$

Thus, (53) and (50) imply that v_t is uncorrelated with lagged values of ξ :

$$E(v_t \xi_\tau') = 0 \qquad \text{for } \tau = t-1, t-2, \ldots, 1. \tag{56}$$

Similarly,

$$E(w_t \xi_\tau') = 0 \qquad \text{for } \tau = 1, 2, \ldots, T \tag{57}$$

$$\begin{aligned} E(w_t, y_\tau') &= E\left[w_t (A' x_\tau + H' \xi_\tau + w_\tau)'\right] \\ &= 0 \qquad \text{for } \tau = t-1, t-2, \ldots, 1 \end{aligned} \tag{58}$$

$$E(v_t y_\tau') = 0 \qquad \text{for } \tau = t-1, t-2, \ldots, 1. \tag{59}$$

The system of (48) through (54) is quite flexible, though it is straightforward to generalize the results further to systems in which v_t is correlated with w_t (see, for example, Anderson and Moore, 1979). The various parameter matrices $(F, Q, A, H, \text{or} R)$ could be functions of time. The presentation will be clearest, however, if we focus on the basic form in (48) through (54).

Examples of State-Space Representations Consider a univariate AR(p) process,

$$\begin{aligned} y_{t+1} - \mu &= \phi_1 (y_t - \mu) + \phi_2 (y_{t-1} - \mu) + \ldots \\ &\quad \phi_p (y_{t-p+1} - \mu) + \epsilon_{t+1}, \\ E(\epsilon_t \epsilon_\tau) &= \begin{cases} \sigma^2 & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{60}$$

This could be written in state-space form as follows.
State Equation $(r = p)$:

$$\begin{bmatrix} y_{t+1}-\mu \\ y_t-\mu \\ \vdots \\ y_{t-p+2}-\mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} y_t-\mu \\ y_{t-1}-\mu \\ \vdots \\ y_{t-p+1}-\mu \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{61}$$

Observation Equation $(n = 1)$:

$$y_t = \mu + \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} y_t-\mu \\ y_{t-1}-\mu \\ \vdots \\ y_{t-p+1}-\mu \end{bmatrix}. \tag{62}$$

That is, we would specify

$$\xi_t = \begin{bmatrix} y_t-\mu \\ y_{t-1}-\mu \\ \vdots \\ y_{t-p+1}-\mu \end{bmatrix} \qquad F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

$$v_{t+1} = \begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad Q = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$y_t = y_t \qquad A' = \mu \qquad x_t = 1$$

$$H' = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \qquad w_t = 0 \qquad R = 0.$$

Note that the state equation here is simply the first-order vector difference equation. The observation equation here is a trivial identity. Thus, we have already seen that the state-space representation (61) and (62) is just another way of summarizing the AR(p) process (60). The reason for rewriting an AR(p) process in such a form was to obtain a convenient summary of the system's dynamics, and this is the basic reason to be interested in the state-space representation of any system.

As another example, consider a univariate MA(1) process,

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}. \tag{63}$$

This could be written in state-space form as follows:
State Equation $(r = 2)$:

$$\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix} \tag{64}$$

Observation Equation $(n = 1)$:

$$y_t = \mu + \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}; \tag{65}$$

that is,

$$\xi_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} \qquad F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \qquad v_{t+1} = \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \qquad y_t = y_t \qquad A' = \mu \qquad x_t = 1$$

$$H' = \begin{bmatrix} 1 & \theta \end{bmatrix} \qquad w_t = 0 \qquad R = 0.$$

There are many ways to write a given system in state-space form. For example, the MA(1) process (63) can also be represented in this way:
State Equation $(r = 2)$:

$$\begin{bmatrix} \epsilon_{t+1} + \theta\epsilon_t \\ \theta\epsilon_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t + \theta\epsilon_{t-1} \\ \theta\epsilon_t \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ \theta\epsilon_{t+1} \end{bmatrix} \tag{66}$$

Observation Equation $(n = 1)$:

$$y_t = \mu + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t + \theta\epsilon_{t-1} \\ \theta\epsilon_t \end{bmatrix}. \tag{67}$$

Note that the original MA(1) representation of (63), the first state-space representation of (64) and (65), and the second state-space representation of (66) and (67) all characterize the same process. We will obtain the identical forecasts of the process or value for the likelihood function from any of the three representation and can feel free to work with whichever is most convenient.

More generally, a univariate ARMA(p, q) process can be written in state-space form by defining $r \equiv \max\{p, q + 1\}$:

$$\begin{aligned} y_t - \mu = {} & \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \cdots + \phi_r(y_{t-r} - \mu) \\ & + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_{r-1}\epsilon_{t-r+1}, \end{aligned} \tag{68}$$

where we interpret $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$. Consider the following state-space representation.
State Equation $(r = \max\{p, q + 1\})$:

$$\xi_{t+1} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{r-1} & \phi_r \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \xi_t + \begin{bmatrix} \epsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{69}$$

Observation Equation $(n = 1)$:

$$y_t = \mu + \begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_{r-1} \end{bmatrix} \xi_t. \tag{70}$$

To verify that (69) and (70) describe the same process as (68), let ξ_{jt} denote the jth element of ξ_t. Thus, the second row of the state equation asserts that

$$\xi_{2,t+1} = \xi_{1t.}$$

The third row asserts that

$$\xi_{3,t+1} = \xi_{2t} = \xi_{1,t-1},$$

and in general the jth row implies that

$$\xi_{j,t+1} = L^{j-1}\xi_{1,t+1}.$$

Thus, the first row of the state equation implies that

$$\xi_{1,t+1} = (\phi_1 + \phi_2 L + \phi_3 L^2 + \cdots + \phi_r L^{r-1})\xi_{1t} + \epsilon_{t+1}$$

or

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)\xi_{1,t+1} = \epsilon_{t+1}. \tag{71}$$

The observation equation states that

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_{r-1} L^{r-1})\xi_{1t}. \tag{72}$$

Multiplying (72) by $(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)$ and using (71) gives

$$\begin{aligned}(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_r L^r)(y_t - \mu) = \\ (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_{r-1} L^{r-1})\epsilon_t,\end{aligned} \tag{73}$$

which indeed reproduces (68).

The state-space form can also be very convenient for modeling sums of stochastic processes or the consequences of measurement error. For example, Stock and Watson (1991) postulated the existence of an unobserved scalar C_t that represents the state of the business cycle. A set of n different observed macroeconomic variables $(y_{1t}, y_{2t}, \ldots, y_{nt})$ are each assumed to be influenced by the business cycle and also to have an idiosyncratic component (denoted χ_{it}) that is unrelated to movements in y_{jt}, for $i \neq j$. If the business cycle and each of the idiosyncratic components could be described by univariate AR(1) processes, then the $[(n + 1) \times 1]$ state vector would be

$$\xi_t = \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix} \tag{74}$$

with state equation

$$\begin{bmatrix} C_{t+1} \\ \chi_{1,t+1} \\ \chi_{2,t+1} \\ \vdots \\ \chi_{n,t+1} \end{bmatrix} = \begin{bmatrix} \phi_C & 0 & 0 & \cdots & 0 \\ 0 & \phi_1 & 0 & \cdots & 0 \\ 0 & 0 & \phi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_n \end{bmatrix} \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix} + \begin{bmatrix} v_{C,t+1} \\ v_{1,t+1} \\ v_{2,t+1} \\ \vdots \\ v_{n,t+1} \end{bmatrix} \tag{75}$$

and observation equation

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \gamma_1 & 1 & 0 & \cdots & 0 \\ \gamma_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_n & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} C_t \\ \chi_{1t} \\ \chi_{2t} \\ \vdots \\ \chi_{nt} \end{bmatrix}. \tag{76}$$

Thus, γ_i is a parameter that describes the sensitivity of the ith series to the business cycle. To allow for pth-order dynamics, Stock and Watson replaced C_t and χ_{it} in (74) with the $p \times 1$ vectors $(C_t, C_{t-1}, \ldots, C_{t-p+1})'$ and $(\chi_{it}, \chi_{i,t-1}, \ldots, \chi_{i,t-p+1})'$ so that $\quad_t$ is an $[(n+1) \times 1]$ vector. The scalars ϕ_i in (75) are then replaced by $(p \times p)$ matrices F_i with the structure of the matrix F in (61), and $[n \times (p-1)]$ blocks of zeros are added between the columns of H' in the observation equation (76).

2.8.2 Derivarion of the Kalman Filter

Overview of the Kalman Filter Consider the general state-space system (48) through (54), whose key equations are reproduced here for convenience:

$$\xi_{t+1} = F\xi_t + v_{t+1} \tag{77}$$

$$y_t = A'x_t + H'\xi_t + w_t \tag{78}$$

$$E(v_t v_\tau') = \begin{cases} Q & \text{for } t = \tau \\ 0 & \text{otherwise} \end{cases} \tag{79}$$

$$E(w_t w_\tau') = \begin{cases} R & \text{for } t = \tau \\ 0 & \text{otherwise.} \end{cases} \tag{80}$$

The analyst is presumed to have observed $y_1, y_2, \ldots, y_T, x_1, x_2, \ldots, x_T$. One of the ultimate objective may be to estimate the values of any unknown parameters in the system on the basis of these observations. For now, however, we will assume that the particular numerical values of F, Q, A, H, and R are known with certainty; Subsection 2.8.4 will give details on how these parameters can be estimated from the data.

There are many uses of the Kalman filter. It is motivated here as an algorithm for calculating linear least squares forecasts of the state vector on the basis on data observed through data t,

$$\hat{\xi}_{t+1|1} \equiv \hat{E}(\xi_{t+1}|\mathcal{Y}_t),$$

where

$$\mathcal{Y}_t \equiv (y_t', y_{t-1}', \ldots, y_1', x_t', x_{t-1}', \ldots, x_1')' \tag{81}$$

and $\hat{E}(\xi_{t+1}|\mathcal{Y}_t)$ denotes the linear projection of ξ_{t+1} on $\mathcal{Y}_t$ and a constant, The Kalman filter calculates these forecasts recursively, generating $\hat{\xi}_{1|0}$, $\hat{\xi}_{2|1}$, $\ldots$, $\hat{\xi}_{T|T-1}$ in succession. Associated with each of these forecasts is a mean squared error (MSE) matrix, represented by the following $(r \times r)$ matrix:

$$P_{t+1|t} \equiv E\left[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'\right]. \tag{82}$$

Starting the Recursion The recursion begins with $\hat{\xi}_{1|0}$, which denotes a forecast of ξ_1 based on no observations of y or x. This is just the unconditional mean of ξ_1,

$$\hat{\xi}_{1|0} = E(\xi_1),$$

with associated MSE

$$P_{1|0} = E\{[\xi_1 - E(\xi_1)]\,[\xi_1 - E(\xi_1)]'\}.$$

For example, for the state-space representation of the MA(1) system given in (64) and (65), the state vector was

$$\xi_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix},$$

for which

$$\hat{\xi}_{1|0} = E\begin{bmatrix} \epsilon_1 \\ \epsilon_0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{83}$$

$$P_{1|0} = E\left(\begin{bmatrix} \epsilon_1 \\ \epsilon_0 \end{bmatrix} \begin{bmatrix} \epsilon_1 & \epsilon_0 \end{bmatrix} \right) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \tag{84}$$

where $\sigma^2 = E(\epsilon_1^2)$.

More generally, if eigenvalues of F are all inside the unit circle, then the process for ξ_t in (77) is covariance-stationary. The unconditional mean of ξ_t can be found by taking expectations of both sides of (77), producing

$$E(\xi_{t+1}) = FE(\xi_t),$$

or, since ξ_t is covariance-stationary,

$$(I_r - F)E(\xi_t) = 0.$$

Since unity is not an eigenvalue of F, the matrix $(I_r - F)$ is not singular, and this equation has the unique solution $E(\xi_t) = 0$. The unconditional variance of ξ can similarly be found by postmultiplying (77) by its transpose and taking expectations:

$$E(\xi_{t+1}\xi'_{t+1}) = E\left[(F\xi_t + v_{t+1})(\xi'_t F' + v'_{t+1})\right] = FE(\xi_t\xi'_t)F' + E(v_{t+1}v'_{t+1}).$$

Cross-product terms have disappeared in light of (56). Letting Σ denote the variance-covariance matrix of ξ, this equation implies

$$\Sigma = F\Sigma F' + Q,$$

whose solution is given by

$$vec(\Sigma) = [I_{r^2} - (F \otimes F)]^{-1} vec(Q).$$

Thus, in general, provided that the eigenvalues of F are inside the unit circle, the Kalman filter iterations can be started with $\hat{\xi}_{1|0} = 0$ and $P_{1|0}$ the $(r \times r)$ matrix whose elements expressed as a column vector are given by

$$vec(P_{1|0}) = [I_{r^2} - (F \otimes F)]^{-1} \cdot vec(Q).$$

If instead some eigenvalues of F are on or outside the unit circle, or if the initial state ξ_1 is not regarded as an arbitrary draw from the process implied by (77), then $\hat{\xi}_{1|0}$ can be replaced with the analyst's best guess as to the initial value of ξ_1, where $P_{1|0}$ is a positive definite matrix summarizing the confidence in this guess. Larger values for the diagonal elements of $P_{1|0}$ register greater uncertainty about the true value of ξ_1.

Forecasting $\mathbf{y_t}$ Given starting values $\hat{\xi}_{1|0}$ and $P_{1|0}$, the next step is to calculate analogous magnitudes for the following date, $\hat{\xi}_{2|1}$ and $P_{t+1|t}$, the goal is to calculate $\hat{\xi}_{t+1|t}$ and $P_{t+1|t}$.

First, note that since we have assumed that x_t contains no information about ξ_t beyond that contained in $\mathcal{Y}_{t-1}$,

$$\hat{E}(\xi_t|x_t, \mathcal{Y}_{t-1}) = \hat{E}(\xi_t|\mathcal{Y}_{t-1}) = \hat{\xi}_{t|t-1}.$$

Next consider forecasting the value of y_t:

$$\hat{y}_{t|t-1} \equiv \hat{E}(y_t|x_t, \mathcal{Y}_{t-1}).$$

Notice from (78) that

$$\hat{E}(y_t|x_t, \xi_t) = A'x_t + H'\xi_t,$$

and so, from the law of iterated projections,

$$\hat{y}_{t|t-1} = A'x_t + H'\hat{E}(\xi_t|x_t, \mathcal{Y}_{t-1}) = A'x_t + H'\hat{\xi}_{t|t-1}. \tag{85}$$

From (78), the error of this forecast is

$$\begin{aligned} y_t - \hat{y}_{t|t-1} &= A'x_t + H'\xi_t + w_t - A'x_t - H'\hat{\xi}_{t|t-1} \\ &= H'(\xi_t - \hat{\xi}_{t|t-1}) + w_t \end{aligned}$$

with MSE

$$\begin{aligned} &E\left[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right] \\ &\qquad = E\left[H'(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'H\right] + E\left[w_t w_t'\right]. \end{aligned} \tag{86}$$

Cross-product terms have disappeared, since

$$E\left[w_t(\xi_t - \hat{\xi}_{t|t-1})'\right] = 0. \tag{87}$$

To justify (87), recall from (57) that w_t is uncorrelated with ξ_t. Furthermore, since $\hat{\xi}_{t|t-1}$ is a linear function of $\mathcal{Y}_{t-1}$, by (58) it too must be uncorrelated with w_t.

Using (80) and (82), equation (86) can be written

$$E\left[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right] = H'P_{t|t-1}H + R. \tag{88}$$

Updating the Inference About ξ_t Next, the inference about the current value of ξ_t is updated on the basis of the observation of y_t to produce

$$\hat{\xi}_{t|t} = \hat{E}(\xi|y_t, x_t, \mathcal{Y}_{t-1}) = \hat{E}(\xi_t|\mathcal{Y}_t).$$

This can be evaluated using the formula for updating a linear projection:

$$\begin{aligned} \hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} &+ \{E\left[(\xi_t - \hat{\xi}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right]\} \\ &\times \{E\left[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right]\}^{-1} \times (y_t - \hat{y}_{t|t-1}). \end{aligned} \tag{89}$$

But

$$\begin{aligned} E\left[(\xi_t - \hat{\xi}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right] &= E\{\left[\xi_t - \hat{\xi}_{t|t-1}\right]\left[H'(\xi_t - \hat{\xi}_{t|t-1}) + w_t\right]'\} \\ &= E\left[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'H\right] \\ &= P_{t|t-1}H \end{aligned} \tag{90}$$

by virtue of (87) and (82). Substituting (90), (88), and (85) into (89) gives

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}). \tag{91}$$

The MSE associated with this updated projection, which is denoted $P_{t|t}$, is as follows:

$$\begin{aligned}
P_{t|t} &\equiv E\left[(\xi_t - \hat{\xi}_{t|t})(\xi_t - \hat{\xi}_{t|t})'\right] \\
&= \left[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'\right] \\
&\quad - \{E\left[(\xi_t - \hat{\xi}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right]\} \\
&\quad \times \{E\left[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'\right]\}^{-1} \\
&\quad \times \{E\left[(y_t - \hat{y}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'\right]\} \\
&= P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1}. \qquad (92)
\end{aligned}$$

Producing a Forecast of ξ_{t+1} Next, the state equation (77) is used to forecast ξ_{t+1} :

$$\begin{aligned}
\hat{\xi}_{t+1|t} &= \hat{E}(\xi_{t+1}|\mathcal{Y}_t) \\
&= F\hat{E}(\xi_t|\mathcal{Y}_t) + \hat{E}(v_{t+1}|\mathcal{Y}_t) \\
&= F\hat{\xi}_{t|t} + 0. \qquad (93)
\end{aligned}$$

Substituting (91) into (93)

$$\begin{aligned}
\hat{\xi}_{t+1|t} &= F\hat{\xi}_{t|t-1} \\
&\quad + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}). \qquad (94)
\end{aligned}$$

The coefficient matrix in (94) is known as the *gain matrix* and is denoted K_t :

$$K_t \equiv FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}, \qquad (95)$$

allowing (94) to be written

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t-1} + K_t(y_t - A'x_t - H'\hat{\xi}_{t|t-1}). \qquad (96)$$

The MSE of this forecast can be found from (93) and the state equation (77):

$$\begin{aligned}
P_{t+1|t} &= E\left[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'\right] \\
&= E\left[(F\xi_t + v_{t+1} - F\hat{\xi}_{t|t})(F\xi_t + v_{t+1} - F\hat{\xi}_{t|t})'\right] \\
&= FE\left[(\xi_t - \hat{\xi}_{t|t})(\xi_t - \hat{\xi}_{t|t})'\right]F' + E\left[v_{t+1}v_{t+1}'\right] \\
&= FP_{t|t}F' + Q, \qquad (97)
\end{aligned}$$

with cross-product terms again clearly zero. Substituting (92) into (97) produces

$$P_{t+1|t} = F\left[P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1}\right]F' + Q. \qquad (98)$$

Summay and Remarks To summarize, the Kalman filer is started with the unconditional mean and variance of ξ_1 :

$$\begin{aligned}\hat{\xi}_{1|0} &= E(\xi_1) \\ P_{1|0} &= E\left\{[\xi_1 - E(\xi_1)]\,[\xi_1 - E(\xi_1)]'\right\}.\end{aligned}$$

Typically, these are given by $\hat{\xi}_{1|0} = 0$ and $vec(P_{1|0}) = [I_{r^2} - (F \otimes F)]^{-1} \cdot vec(Q)$, but in this paper the calculation of initial conditions are performed in line with Casals, Jerez and Sotoca (2000). We then iterate on

$$\begin{aligned}\hat{\xi}_{t+1|t} &= F\hat{\xi}_{t|t-1} \\ &\quad + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1})\end{aligned} \tag{99}$$

and (98) for $t = 1, 2, \ldots, T$. The value $\hat{\xi}_{t+1|t}$ denotes the best forecast of ξ_{t+1} based on a constant and a linear function of $(y_t, y_{t-1}, \ldots, y_1, x_t, x_{t-1}, \ldots, x_1)$. The matrix $P_{t+1|t}$ gives the MSE of this forecast. The forecast of y_{t+1} is given by

$$\hat{y}_{t+1|t} \equiv \hat{E}(y_{t+1}|x_{t+1}, \mathcal{Y}_t) = A'x_{t+1} + H'\hat{\xi}_{t+1|t} \tag{100}$$

with associated MSE

$$E\left[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})'\right] = H'P_{t+1|t}H + R. \tag{101}$$

It is worth noting that the recursion in (98) could be calculated without ever evaluating (99). The values for $P_{t|t-1}$ in (98) and K_t in (95) are not functions of data, but instead are determined entirely by the population parameters of the process.

An alternative way of writing the recursion for $P_{t+1|t}$ is sometimes useful. Subtracting the Kalman updating equation (96) from the state equation (77) produces

$$\xi_{t+1} - \hat{\xi}_{t+1|t} = F(\xi_t - \hat{\xi}_{t|t-1}) - K_t(y_t - A'x_t - H'\hat{\xi}_{t|t-1}) + v_{t+1}. \tag{102}$$

Further substituting the observation equation (78) into (102) results in

$$\xi_{t+1} - \hat{\xi}_{t+1|t} = (F - K_tH')(\xi_t - \hat{\xi}_{t|t-1}) - K_tw_t + v_{t+1}. \tag{103}$$

Postmultiplying (103) by its transpose and taking expectations,

$$\begin{aligned}&E\left[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'\right] \\ &\quad = (F - K_tH')E\left[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})'\right](F' - HK_t') + K_tRK_t' + Q;\end{aligned}$$

or, recalling the definition of $P_{t+1|t}$ in equation (82),

$$P_{t+1|t} = (F - K_tH')P_{t|t-1}(F' - HK_t') + K_tRK_t' + Q. \tag{104}$$

Equation (104) along with the definition of K_t in (95) will produce the same sequence generated by equation (98).

2.8.3 Forecasts Based on the State-Space Representation

The Kalman filter computations in (98) through (101) are normally calculated by computer, using the known numerical values of F, Q, A, H, and R along with the actual data. To help make the ideas more concrete, however, we now explore analytically the outcome of these calculations for a simple example.

Example - Using the Kalman Filter to Find Exact Finite-Sample Forecasts for an MA(1) Process Consider again a state-space representation for the MA(1) process:
State Equation $(r = 2)$:

$$\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix} \tag{105}$$

Observation Equation $(n = 1)$:

$$y_t = \mu + \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} \tag{106}$$

$$\xi_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} \tag{107}$$

$$F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{108}$$

$$v_{t+1} = \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix} \tag{109}$$

$$Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{110}$$

$$y_t = y_t \tag{111}$$

$$A' = \mu \tag{112}$$

$$x_t = 1 \tag{113}$$

$$H' = \begin{bmatrix} 1 & \theta \end{bmatrix} \tag{114}$$

$$w_t = 0 \tag{115}$$

$$R = 0. \tag{116}$$

The starting values for the filter were described in (83) and (84):

$$\hat{\xi}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$P_{1|0} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}.$$

Thus, from (100), the period 1 forecast is

$$\hat{y}_{1|0} = \mu + H'\hat{\xi}_{1|0} = \mu,$$

with MSE given by (101):

$$E(y_1 - \hat{y}_{1|0})^2 = H'P_{1|0}H + R = \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix} + 0 = \sigma^2(1 + \theta^2).$$

These, of course, are just the unconditional mean and variance of y.

To see the structure of the recursion of $t = 2, 3, \ldots, T$, consider the basic form of the updating equation (99). Notice that since the first row of F consists entirely of zeros, the first element of the vector $\hat{\xi}_{t+1|t}$ will always equal to zero, for all t. We see why if we recall the meaning of the state vector in (107):

$$\hat{\xi}_{t+1|t} = \begin{bmatrix} \hat{\epsilon}_{t+1|t} \\ \hat{\epsilon}_{t|t} \end{bmatrix}. \tag{117}$$

Naturally, the forecast of the future white noise, $\hat{\epsilon}_{t+1|t}$, is always zero. The forecast of y_{t+1} is given by (100):

$$\hat{y}_{t+1|t} = \mu + \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_{t+1|t} \\ \hat{\epsilon}_{t|t} \end{bmatrix} = \mu + \theta\hat{\epsilon}_{t|t}. \tag{118}$$

The Kalman filter updating equation for the MSE, equation (97), for this example becomes

$$P_{t+1|t} = FP_{t|t}F' + Q = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} P_{t|t} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}. \tag{119}$$

Thus, $P_{t+1|t}$ is a diagonal matrix of the form

$$P_{t+1|t} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_{t+1} \end{bmatrix}, \tag{120}$$

where the $(2, 2)$ element of $P_{t+1|t}$ (which we have denoted by p_{t+1}) is the same as the $(1, 1)$ element of $P_{t|t}$. Recalling (82) and (117), this term has the interpretation as the MSE of $\hat{\epsilon}_{t|t}$:

$$p_{t+1} = E(\epsilon_t - \hat{\epsilon}_{t|t})^2. \tag{121}$$

The $(1, 1)$ element of $P_{t+1|t}$ has the interpretation as the MSE of $\hat{\epsilon}_{t+1|t}$. We have seen that this forecast is always zero, and its MSE in (120) is σ^2 for all t. The fact that $P_{t+1|t}$ is a diagonal matrix means that the forecast error $(\epsilon_{t+1} - \hat{\epsilon}_{t+1|t})$ is uncorrelated with $(\epsilon_t - \hat{\epsilon}_{t|t})$.

The MSE of the forecast of y_{t+1} is given by (101):

$$\begin{aligned} E(y_{t+1} - \hat{y}_{t+1|t})^2 &= H' P_{t+1|t} H + R \\ &= \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_{t+1} \end{bmatrix} + 0 \\ &= \sigma^2 + \theta^2 p_{t+1}. \end{aligned} \tag{122}$$

Again, the intuition can be seen from the nature of the forecast in (118):

$$\begin{aligned} E(y_{t+1} - \hat{y}_{t+1|t})^2 &= E\left[(\mu + \epsilon_{t+1} + \theta\epsilon_t) - (\mu + \theta\hat{\epsilon}_{t|t})\right]^2 \\ &= E(\epsilon_{t+1}^2) + \theta^2 E(\epsilon_t - \hat{\epsilon}_{t|t})^2, \end{aligned}$$

which, from (121), reproduces (122).

From (99), the series for $\hat{\epsilon}_{t|t}$ is generated recursively from

$$\begin{aligned} \begin{bmatrix} 0 \\ \hat{\epsilon}_{t|t} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \hat{\epsilon}_{t-1|t-1} \end{bmatrix} \\ &\quad + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix} \left\{1/\left[\sigma^2 + \theta^2 p_t\right]\right\} \cdot \left\{y_t - \mu - \theta\hat{\epsilon}_{t-1|t-1}\right\} \end{aligned}$$

or

$$\hat{\epsilon}_{t|t} = \left\{\sigma^2/\left[\sigma^2 + \theta^2 p_t\right]\right\} \cdot \left\{y_t - \mu - \theta\hat{\epsilon}_{t-1|t-1}\right\} \tag{123}$$

starting from the initial value $\hat{\epsilon}_{0|0} = 0$.

The gain matrix K_t in equation (95) is given by

$$K_t = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix} \left(\frac{1}{\sigma^2 + \theta^2 p_t}\right) = \begin{bmatrix} 0 \\ \sigma^2/\left[\sigma^2 + \theta^2 p_t\right] \end{bmatrix}. \tag{124}$$

Finally, notice from (92) that

$$P_{t|t} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix} - \left(\frac{1}{\sigma^2 + \theta^2 p_t}\right) \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix} \begin{bmatrix} 1 \\ \theta \end{bmatrix} \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \sigma^2 & 0 \\ 0 & p_t \end{bmatrix}.$$

The $(1,1)$ element of $P_{t|t}$ (which we saw equals p_{t+1}) is thus given by

$$p_{t+1} = \sigma^2 - \left\{1/\left[\sigma^2 + \theta^2 p_t\right]\right\} \cdot \sigma^4 = \frac{\sigma^2 \theta^2 p_t}{\sigma^2 + \theta^2 p_t}. \tag{125}$$

The recursion in (125) is started with $p_1 = \sigma^2$ and thus has the solution

$$p_{t+1} = \frac{\sigma^2 \theta^{2t}}{1 + \theta^2 + \theta^4 + \ldots + \theta^{2t}}. \tag{126}$$

It is interesting to note what happens to the filter as t becomes large. First consider the case when $|\theta| \leq 1$. Then, from (126),

$$\lim_{t \to \infty} p_{t+1} = 0$$

and so, from (121),

$$\hat{\epsilon}_{t|t} \xrightarrow{p} \epsilon_t.$$

Thus, given a sufficient number of observations on y, the Kalman filer inference $\hat{\epsilon}_{t|t}$ converges to the true value ϵ_t, and the forecast (118) converges to that of the Wold representation for the process. The Kalman gain in (118) converges to $(0, 1)'$.

Alternatively, consider the case when $|\theta| > 1$. From (126), we have

$$p_{t+1} = \frac{\sigma^2\theta^{2t}(1-\theta^2)}{1-\theta^{2(t+1)}} = \frac{\sigma^2(1-\theta^2)}{\theta^{-2t}-\theta^2}$$

and

$$\lim_{t\to\infty} p_{t+1} = \frac{\sigma^2(1-\theta^2)}{-\theta^2} > 0.$$

No matter how many observations are obtained, it will not be possible to know with certainty the value of the nonfundamental innovation ϵ_t associated with date t on the basis of $(y_t, y_{t-1}, \ldots, y_1)$. The gain is given by

$$\frac{\sigma^2}{\sigma^2+\theta^2 p_t} \to \frac{\sigma^2}{\sigma^2-\sigma^2(1-\theta^2)} = \frac{1}{\theta^2},$$

and the recursion (123) approaches

$$\hat{\epsilon}_{t|t} = (1/\theta^2)(y_t - \mu - \theta\hat{\epsilon}_{t-1|t-1})$$

or

$$\theta\hat{\epsilon}_{t|t} = (1/\theta)(y_t - \mu - \theta\hat{\epsilon}_{t-1|t-1}).$$

Recalling (118), we thus have

$$\hat{y}_{t+1|t} - \mu = (1/\theta)\left[(y_t - \mu) - (\hat{y}_{t|t-1} - \mu)\right]$$

or

$$\hat{y}_{t+1|t} - \mu = (1/\theta)(y_t - \mu) - (1/\theta)^2(y_{t-1} - \mu) + (1/\theta)^3(y_{t-2} - \mu) - \ldots,$$

which again is the AR(∞) forecast associated with the invertible MA(1) representation. Indeed, the forecast of the Kalman filter with θ replaced by θ^{-1} and σ^2 replaced by $\theta^2\sigma^2$ will be identical for any t.

Calculating s-Period-Ahead Forecasts with the Kalman Filter The forecast of y_t calculated in (100) is an exact finite-sample forecast of y_t on the basis of x_t and $\mathcal{Y}_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)$. If x_t is deterministic, it is also easy to use the Kalman filter to calculate exact finite-sample s-period-ahead forecasts.

The state equation (77) can be solved by recursive substitution to yield

$$\xi_{t+s} = F^s\xi_t + F^{s-1}v_{t+1} + F^{s-2}v_{t+2} + \ldots + F^1 v_{t+s-1} + v_{t+s} \quad \text{for } s = 1, 2, \ldots. \tag{127}$$

The projection of ξ_{t+s} on ξ_t and $\mathcal{Y}_t$ is given by

$$\hat{E}(\xi_{t+s}|\xi_t, \mathcal{Y}_t) = F^s\xi_t. \tag{128}$$

Thus, from (127) the s-period-ahead forecast error for the state vector is

$$\begin{aligned}\xi_{t+s} - \hat{\xi}_{t+s|t} =& F^s(\xi_t - \hat{\xi}_{t|t}) + F^{s-1}v_{t+1} + F^{s-2}v_{t+2} \\ &+ \ldots + F^1 v_{t+s-1} + v_{t+s}\end{aligned} \tag{129}$$

with MSE

$$\begin{aligned}P_{t+s|t} =& F^s P_{t|t}(F')^s + F^{s-1}Q(F')^{s-1} + F^{s-2}Q(F')^{s-2} \\ &+ \ldots + FQF' + Q.\end{aligned} \tag{130}$$

To forecast the observed vector y_{t+s}, recall from the observation equation that

$$y_{t+s} = A'x_{t+s} + H'\xi_{t+s} + w_{t+s}. \tag{131}$$

There are advantages if the state vector is defined in such a way that x_t is deterministic, so that the dynamics of any exogenous variables can be represented through ξ_t. If x_t is deterministic, the s-period-ahead forecast of y is

$$\hat{y}_{t+s|t} \equiv \hat{E}(y_{t+s|\mathcal{Y}_t}) = A'x_{t+s} + H'\hat{\xi}_{t+s|t}. \tag{132}$$

The forecast error is

$$\begin{aligned}y_{t+s} - \hat{y}_{t+s|t} &= (A'x_{t+s} + H'\xi_{t+s} + w_{t+s}) - (A'x_{t+s} + H'\hat{\xi}_{t+s|t}) \\ &= H'(\xi_{t+s} - \hat{\xi}_{t+s|t}) + w_{t+s}\end{aligned}$$

with MSE

$$E[(y_{t+s} - \hat{y}_{t+s|t})(y_{t+s} - \hat{y}_{t+s|t})'] = H'P_{t+s|t}H + R. \tag{133}$$

2.8.4 Maximum Likelihood Estimation of Parameters

Using the Kalman Filter to Evaluate the Likelihood Function The Kalman filter was motivated in Subsection 2.8.2 in terms of linear projections. The forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ are thus optimal within the set of forecasts that are linear in $(x_t, \mathcal{Y}_{t-1})$. If the initial state ξ_1 and the innovations $\{w_t, v_t\}_{t=1}^T$ are multivariate Gaussian, then we make the stronger claim that the forecasts $\hat{\xi}_{t|t-1}$ and $\hat{y}_{t|t-1}$ calculated by the Kalman filter are optimal among any functions of $(x_t, \mathcal{Y}_{t-1})$. Moreover, if ξ_1 and $\{w_t, v_t\}_{t=1}^T$ are

Gaussian, then the distribution of y_t conditional on $(x_t, \mathcal{Y}_{t-1})$ is Gaussian with mean given by (100) and variance given by (101):

$$y_t|x_t, \mathcal{Y}_{t-1} \sim N\left((A'x_t + H'\hat{\xi}_{t|t-1}), (H'P_{t|t-1}H + R)\right);$$

that is,

$$\begin{aligned} f_{Y_t,X_t,\mathcal{Y}_{t-1}}&(y_t|x_t, \mathcal{Y}_{t-1}) \\ &= (2\pi)^{-n/2}|H'P_{t|t-1}H + R|^{-1/2} \\ &\quad \times \exp\{-\tfrac{1}{2}(y_t - A'x_t - H'\hat{\xi}_{t|t-1})'(H'P_{t|t-1}H + R)^{-1} \\ &\quad \times (y_t - A'x_t - H'\hat{\xi}_{t|t-1})\} \qquad \text{for } t = 1, 2, \ldots, T. \end{aligned} \tag{134}$$

From (134), it is a simple matter to construct the sample log likelihood,

$$\sum_{t=1}^{T} \log f_{Y_t,X_t,\mathcal{Y}_{t-1}}(y_t|x_t, \mathcal{Y}_{t-1}). \tag{135}$$

Expression (135) can then be maximized numerically with respect to the unknown parameters in the matrices F, Q, A, H, and R; see Burmeister and Wall (1982) for an illustrative application.

As stressed by Harvey and Phillips (1979), this representation of the likelihood is paticularly convenient for estimating regressions involving moving average terms. Moreover, (135) gives the exact log likelihood function, regardless of whether the moving average representation is invertible.

As an illustrative example, suppose we wanted to estimate a bivariate regression model whose equations were

$$\begin{aligned} y_{1t} &= a_1'x_t + u_{1t} \\ y_{2t} &= a_2'x_t + u_{2t}, \end{aligned}$$

where x_t is a $(k \times 1)$ vector of exogenous explanatory variables and a_1 and a_2 are $(k \times 1)$ vectors of coefficients; if the two regressions have different explanatory variables, the variables from both regressions are included in x_t with zeros appropriately imposed on a_1 and a_2. Suppose that the disturbance vector follows a bivariate MA(1) process:

$$\begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1} \\ \epsilon_{2.t-1} \end{bmatrix},$$

with $(\epsilon_{1t}, \epsilon 2t)' \sim$ i.i.d. $N(0, \Omega)$. This model can be written in state-space

form by defining

$$\xi = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{bmatrix} \qquad F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \qquad v_{t+1} = \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \\ 0 \\ 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad A' = \begin{bmatrix} a_1' \\ a_2' \end{bmatrix}$$

$$H' = \begin{bmatrix} 1 & 0 & \theta_{11} & \theta 12 \\ 0 & 1 & \theta_{21} & \theta 22 \end{bmatrix} \qquad R = 0,$$

where $\sigma_{ij} = E(\epsilon_{it}\epsilon_{jt})$. The Kalman filter iteration is started from

$$\hat{\xi}_{1|0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad P_{1|0} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{21} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{11} & \sigma_{12} \\ 0 & 0 & \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

Maximization of (135) is started by making an initial guess as to the numerical values of the unknown parameters. One obvious way to do this is to regress y_{1t} on the elements of x_t that appear in the first equation to get an initial guess for a_1. A similar Ordinary Least Squares (OLS) regression for y_{2t} yields a guess for a_2. Setting $\theta_{11} = \theta_{12} = \theta_{21} = \theta_{22} = 0$ initially, a first guess for Ω could be the estimated variance-covariance matrix of the residuals from these two OLS regressions. For these initial numerical values for the population parameters, we could construct F, Q, A, H, and R from the expressions just given and iterate on (98) through (101) for $t = 1, 2, \ldots,$ $T - 1$. The sequences $\{\hat{\xi}_{t|t-1}\}_{t=1}^{T}$ and $\{P_{t|t-1}\}_{t=1}^{T}$ resulting from these iterations could then be used in (134) and (135) to calculate the value for the log likelihood function that results from these initial parameter values. The numerical optimization methods can then be employed to make better guesses as to the value of the unknown parameters until (135) is maximized. It can be shown (see Hamilton, 1994, Section 5.9) that the numerical search will be better behaved if Ω is parameterized in terms of its Cholesky factorization.

As a second example, consider a scalar Gaussian ARMA$(1, 1)$ process,

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t + \theta\epsilon_{t-1},$$

with $\epsilon_t \sim$i.i.d. $N(0, \sigma^2)$. This can be written in state-space form as in (69)

and (70) with $r = 2$ and

$$F = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \qquad v_{t+1} = \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix} \qquad Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$$
$$A' = \mu \qquad x_t = 1 \qquad H' = \begin{bmatrix} 1 & \theta \end{bmatrix} \qquad R = 0$$
$$\hat{\xi}_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad P_{1|0} = \begin{bmatrix} \sigma^2/(1-\phi^2) & \phi\sigma^2/(1-\phi^2) \\ \phi\sigma^2/(1-\phi^2) & \sigma^2/(1-\phi^2) \end{bmatrix}.$$

This value for $P_{1|0}$ was obtained by recognizing that the state equation (69) describes the behavior of $\xi = (z_t, z_{t-1}, \ldots, z_{t-r+1})'$, where $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_r z_{t-r} + \epsilon_t$ follows an $AR(r)$ process. For this example, $r = 2$, so that $P_{1|0}$ is the variance-covariance matrix of two consecutive draws from an $AR(2)$ process with parameters $\phi_1 = \phi$ and $\phi_2 = 0$. The expressions just given for F, Q, A, H, and R are then used in the Kalman filter iterations. Thus, expression (135) allows easy computation of the exact likelihood function for an ARMA(p, q) process. This computation is valid regardless of whether the moving average parameters satisfy the invertibility condition. Similarly, expression (132) gives the exact finite-sample s-period-ahead forecast for the process and (133) its MSE, again regardless of whether the invertible representation is used.

Typically, numerical search procedures for maximizing (135) require the derivatives of the log likelihood. These can be calculated numerically or analytically. To characterize the analytical derivatives of (135), collect the unknown parameters to be estimated in a vector θ, and write $F(\theta)$, $Q(\theta)$, $A(\theta)$, $H(\theta)$, and $R(\theta)$. Implicitly, then, $\hat{\epsilon}_{t|t-1}(\theta)$ and $P_{t|t-1}(\theta)$ will be functions of θ as well, and the derivative of the log of (134) with respect to the ith element of θ will involve $\partial\hat{\xi}_{t|t-1}(\theta)/\partial\theta_i$ and $\partial P_{t|t-1}(\theta)/\partial\theta_i$. These derivatives can also be generated recursively by differentiating the Kalman filter recursion, (98) and (99), with respect to θ_i; see Caines (1988, pp.585-86) for illustration.

For many state-space models, the EM algorithm of Dempster, Laird, and Rubin (1977) offers a particularly convenient means for maximizing (135), as developed by Shumway and Stoffer (1982) and Watson and Engle (1983).

Identification Although the state-space representation gives a very convenient way to calculate the exact likelihood function, a word of caution should be given. In the absence of restrictions on F, Q, A, H, and R, the parameters of the state-space representation are unidentified - more than one set of values for the parameters can give rise to the identical value of the likelihood function, and the data give us no guide for choosing among these. A trivial example is the following system:
State Equation ($r = 2$):

$$\xi_{t+1} = \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{bmatrix} \tag{136}$$

Observation Equation $(n = 1)$:

$$y_t = \epsilon_{1t} + \epsilon_{2t}. \tag{137}$$

Here, $F = 0$, $Q = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, $A' = 0$, $H' = \begin{bmatrix} 1 & 1 \end{bmatrix}$, and $R = 0$. This model asserts that y_t is white noise, with mean zero and variance given by $(\sigma_1^2 + \sigma_2^2)$. The log likelihood function from (134) and (135) simplifies to

$$\begin{aligned} &\log f_{Y_T, Y_{T-1}, \ldots, Y_1}(y_T, y_{T-1}, \ldots, y_1) \\ &\quad = -(T/2)\log(2\pi) - (T/2)\log(\sigma_1^2 + \sigma_2^2) - \sum_{t=1}^{T} y_t^2/[2(\sigma_1^2 + \sigma_2^2)]. \end{aligned} \tag{138}$$

Clearly, any values for σ_1^2 and σ_2^2 that sum to a given constant will produce identical value for the likelihood function.

The $MA(1)$ process explored in Subsection 2.8.3 provides a second example of an unidentified state-space representation. As it can be verified, the identical value for the log likelihood function (135) would result if θ is replaced by θ^{-1} and σ^2 by $\theta^2\sigma^2$.

These two examples illustrate two basic forms in which absence of identification can occur. Following Rothenberg (1971), a model is said to be *globally identified* at a particular parameter value θ_0 if for any value of θ there exists a possible realization $\mathcal{Y}_T$ for which the value of the likelihood at θ is different from the value of the likelihood at θ_0. A model is said to be *locally identified* at θ_0 if there exists a $\delta > 0$ such that for any value of θ satisfying $(\theta - \theta_0)(\theta - \theta_0)' < \delta$, there exists a possible realization of $\mathcal{Y}_T$ for which the value of the likelihood at θ is different from the value of the likelihood at θ_0. Thus, global identification implies local identification. The first example, (136) and (137), is neither globally nor locally identified, while the MA(1) example is locally identified but globally unidentified.

Local identification is much easier to test for than global identification. Rothenberg (1971) showed that a model is locally identified at θ_0 if and only if the information matrix is nonsingular in a neighborhood around θ_0. Thus, a common symptom of trying to estimate an unidentified model is difficulty with inverting the matrix of second derivatives of the log likelihood function. One approach to checking for local identification is to translate the state-space representation back into a vector ARMA model and check for satisfaction of the conditions in Hannan (1971); see Hamilton (1985) for an example of this approach. A second approach is to work directly with state-space representation, as is done in Gevers and Wertz (1984) and Wall (1987). For an illustration of the second approach, see Burmeister, Wall, and Hamilton (1986).

Asymptotic Properties of Maximum Likelihood Estimates If certain regularity conditions are satisfied, then Caines (1988, Chapter 7) showed

that the maximum likelihood estimate $\hat{\theta}_T$ based on a sample of size T is consistent and asymptotically normal. These conditions include the following: (1) the model must be identified; (2) eigenvalues of F are inside the unit circle; (3) apart from a constant term, the variables x_t behave asymptotically like a full-rank linearly indeterministic covariance-stationary process; and (4) the true value of θ does not fall on a boundary of the allowable parameter space. Pagan (1980, Theorem 4) and Ghosh (1989) examined special cases of state-space models for which

$$\sqrt{T}\mathscr{I}_{2D,T}^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{L} N(0, I_a), \tag{139}$$

where a is the number of elements of θ and $\mathscr{I}_{2D,T}$ is a $(a \times a)$ information matrix for a sample of size T as calculated from a second derivatives of the log likelihood function:

$$\mathscr{I}_{2D,T} = -\frac{1}{T}E\left[\sum_{t=1}^{T} \frac{\partial^2 \log f(y_t|x_t, \mathcal{Y}_{t-1}; \theta)}{\partial\theta\, \partial\theta'}\bigg|_{\theta=\theta_0}\right]. \tag{140}$$

A common practice is to assume that the limit of $\mathscr{I}_{2D,T}$ as $T \to \infty$ is the same as the plim of

$$\hat{\mathscr{I}}_{2D,T} = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 \log f(y_t|x_t, \mathcal{Y}_{t-1}; \theta)}{\partial\theta\, \partial\theta'}\bigg|_{\theta=\hat{\theta}_T}, \tag{141}$$

which can be calculated analytically or numerically by differencing (135). Reported standard errors for $\hat{\theta}_T$ are then square roots of diagonal elements of $(1/T)(\hat{\mathscr{I}}_{2D,T})^{-1}$.

Quasi-Maximum Likelihood Estimation Even if the disturbances v_t and w_t are non-Gaussian, the Kalman filter can still be used to calculate the linear projection of y_{t+s} on past observables. Moreover, we can form the function (135) and maximize it with respect to θ even for non-Gaussian systems. This procedure will still yield consistent and asymptotically Normal estimates of the elements of F, Q, A, H, and R. Watson (1989, Theorem 2) presented conditions under which the quasi-maximum likelihood estimates satisfy

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{L} N(0, [\mathscr{I}_{2D}\mathscr{I}_{OP}^{-1}\mathscr{I}_{2D}]^{-1}), \tag{142}$$

where $\mathscr{I}_{2D}$ is the plim of (141) when evaluated at the true value θ_0 and $\mathscr{I}_{OP}$ is the plim of the outer-product estimate of the information matrix,

$$\mathscr{I}_{OP} = \text{plim } (1/T)\sum_{t=1}^{T}[h(\theta_0, \mathcal{Y}_t)][h(\theta_0, \mathcal{Y}_t)]',$$

where

$$h(\theta_0, \mathcal{Y}_t) \equiv \frac{\partial \log f(y_t|x_t, \mathcal{Y}_{t-1}; \theta)}{\partial\theta}\bigg|_{\theta=\theta_0}.$$

2.8.5 Smoothing

The Kalman filter was motivated in Subsection 2.8.2 as an algorithm for calculating a forecast of the state vector ξ_t as a linear function of previous observations,

$$\hat{\xi}_{t|t-1} \equiv \hat{E}(\xi_t|\mathcal{Y}_{t-1}), \tag{143}$$

where $\mathcal{Y}_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots, y'_1, x'_{t-1}, x'_{t-2}, \ldots, x'_1)'$. The matrix $P_{t|t-1}$ represented the MSE of this forecast:

$$P_{t|t-1} \equiv E[(\xi_t - \hat{\xi}_{t|t-1})(\xi_t - \hat{\xi}_{t|t-1})']. \tag{144}$$

For many uses of the Kalman filter these are the natural magnitudes of interest. In some settings, however, the state vector ξ_t is given a structural interpretation, in which case the value of this unobserved variable might be of interest for its own sake. For example, in the model of the business cycle by Stock and Watson, it would be helpful to know the state of the business cycle at any historical date t. A goal might then be to form an inference about the value of ξ_t based on the full set of data collected, including observations of $y_t, y_{t+1}, \ldots, y_T, x_t, x_{t+1}, \ldots, x_T$. Such an inference is called the *smoothed* estimate of ξ_t, denoted

$$\hat{\xi}_{t|T} \equiv \hat{E}(\xi_t|\mathcal{T}). \tag{145}$$

For example, data on GNP from 1954 through 1990 might be used to estimate the value that ξ took on in 1960. The MSE of this smoothed estimate is denoted

$$P_{t|T} \equiv E[(\xi_t - \hat{\xi}_{t|T})(\xi_t - \hat{\xi}_{t|T})']. \tag{146}$$

In general, $P_{t|\tau}$ denotes the MSE of an estimate of ξ_t that is based on observations of y and x through date τ.

For the reader's convenience, we reproduce here the key equations for the Kalman filter:

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\hat{\xi}_{t|t-1}) \tag{147}$$

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t} \tag{148}$$

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1} \tag{149}$$

$$P_{t+1|t} = FP_{t|t}F' + Q. \tag{150}$$

Consider the estimate of ξ_t based on observations through date t, $\hat{\xi}_{t|t}$. Suppose we were subsequently told the true value of ξ_{t+1}. From the formula for updating a linear projection, the new estimate of ξ_t could be expressed as

$$\begin{aligned}\hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_t) = \hat{\xi}_{t|t} &+ \{E[(\xi_t - \hat{\xi}_{t|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})']\} \\ &\times \{E[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})']\}^{-1} \\ &\times (\xi_{t+1} - \hat{\xi}_{t+1|t}).\end{aligned} \tag{151}$$

The first term in the product on the right side of (151) can be written

$$E[(\xi_t - \hat{\xi}_{t|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'] = E[(\xi_t - \hat{\xi}_{t|t})(F\xi_t + v_{t+1} - F\hat{\xi}_{t|t})'],$$

by virtue of (77) and (148). Furthermore, v_{t+1} is uncorrelated with ξ and $\hat{\xi}_{t|t}$. Thus,

$$E[(\xi_t - \hat{\xi}_{t|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})'] = E[(\xi_t - \hat{\xi}_{t|t})(\xi_t - \hat{\xi}_{t|t})'F'] = P_{t|t}F'. \quad (152)$$

Substituting (152) and the definition of $P_{t+1|t}$ into (151) produces

$$\hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_t) = \hat{\xi}_{t|t} + P_{t|t}F'P_{t+1|t}^{-1}(\xi_{t+1} - \hat{\xi}_{t+1|t}.)$$

Defining

$$J_t \equiv P_{t|t}F'P_{t+1|t}^{-1}, \quad (153)$$

we have

$$\hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_t) = \hat{\xi}_{t|t} + J_t(\xi_{t+1} - \hat{\xi}_{t+1|t}). \quad (154)$$

Now, the linear projection in (154) turns out to be the same as

$$\hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_T); \quad (155)$$

that is, knowledge of y_{t+j} or x_{t+j} for $j > 0$ would be of no added value if we already knew the value of ξ_{t+1}. To see this, note that y_{t+j} can be written as

$$y_{t+j} = A'x_{t+j} + H'(F^{j-1}\xi_{t+1} + F^{j-2}v_{t+2} + F^{j-3}v_{t+3} + \ldots + v_{t+j}) + w_{t+j},$$

But the error

$$\xi_t - \hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_t) \quad (156)$$

is uncorrelated with ξ_{t+1}, by the definition of a linear projection, and uncorrelated with x_{t+j}, w_{t+j}, v_{t+j}, v_{t+j-1}, $\ldots$, v_{t+2} under the maintained assumptions. Thus, the error (156) is uncorrelated with y_{t+j} or x_{t+j} for $j > 0$, meaning that (155) and (154) are the same, as claimed:

$$\hat{E}(\xi_t|\xi_{t+1}, \mathcal{Y}_T) = \hat{\xi}_{t|t} + J_t(\xi_{t+1} - \hat{\xi}_{t+1|t}). \quad (157)$$

It follows from the law of iterated projections that the smoothed estimate, $\hat{E}(\xi_t|\mathcal{Y}_T)$, can be obtained by projecting (157) on $\mathcal{Y}_T$. In calculating this projection, we need to think carefully about the nature of the magnitudes in (157). The first term, $\hat{\xi}_{t|t}$, indicates a particular *exact* linear function of $\mathcal{Y}_t$; the coefficients of this function are constructed from population moments, and these coefficients should be viewed as deterministic constants from the point of view of performing a subsequent projection. The projection of $\hat{\xi}_{t|t}$ on $\mathcal{Y}_T$ is thus still $\hat{\xi}_{t|t}$, this same linear function of $\mathcal{Y}_t$ - we

can't improve on a perfect fit.[1] The term J_t in (153) is also a function of population moments, and so is again treated as deterministic for purposes of any linear projection. The term $\hat{\xi}_{t+1|t}$ is another exact linear function of $\mathcal{Y}_t$. Thus, projecting (157) on $\mathcal{Y}_T$ turns out to be trivial:

$$\hat{E}(\xi_t|\mathcal{Y}_T) = \hat{\xi}_{t|t} + J_t[\hat{E}(\xi_{t+1}|\mathcal{Y}_T) - \hat{\xi}_{t+1|t}],$$

or

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} + J_t(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}) \tag{158}$$

Thus, the sequence of smoothed estimates $\{\hat{\xi}_{t|T}\}_{t=1}^T$ is calculated as follows. First, the Kalman filter, (147) to (150), is calculated and the sequences $\{\hat{\xi}_{t|t}\}_{t=1}^T$, $\{\hat{\xi}_{t+1|t}\}_{t=0}^{T-1}$, $\{P_{t|t}\}_{t=1}^T$, $\{P_{t+1|t}\}_{t=0}^{T-1}$ are stored. The smoothed estimate for the final date in the sample $\hat{\xi}_{T|T}$, is just the last entry in $\{\hat{\xi}_{t|t}\}_{t=1}^T$. Next, (153) is used to generate $\{J_t\}_{t=1}^{T-1}$. From this, (158) is used for $t = T-1$ to calculate

$$\hat{\xi}_{T-1|T} = \hat{\xi}_{T-1|T-1} + J_{T-1}(\hat{\xi}_{T|T} - \hat{\xi}_{T|T-1}).$$

Now that $\hat{\xi}_{T-1|T}$ has been calculated, (158) can be used for $t = T-2$ to evaluate

$$\hat{\xi}_{T-2|T} = \hat{\xi}_{T-2|T-2} + J_{T-2}(\hat{\xi}_{T-1|T} - \hat{\xi}_{T-1|T-2}).$$

Proceeding backward through the sample in this fashion permits calculation of the full set of smoothed estimates, $\{\hat{\xi}_{t|T}\}_{t=1}^T$.

Next, consider the mean squared error associated with the smoothed estimate. Subtracting both sides of (158) from ξ_t produces

$$\xi_t - \hat{\xi}_{t|T} = \xi_t - \hat{\xi}_{t|t} - J_t\hat{\xi}_{t+1|T} + J_t\hat{\xi}_{t+1|t}$$

or

$$\xi_t - \hat{\xi}_{t|T} + J_t\hat{\xi}_{t+1|T} = \xi_t - \hat{\xi}_{t|t} + J_t\hat{\xi}_{t+1|t}.$$

Multiplying this equation by its transpose and taking expectations,

$$\begin{aligned} &E[(\xi_t - \hat{\xi}_{t|T})(\xi_t - \hat{\xi}_{t|T})'] + J_tE[(\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T})]J'_t \\ &\qquad = E[(\xi_t - \hat{\xi}_{t|t})(\xi_t - \hat{\xi}_{t|t})'] + J_tE[(\hat{\xi}_{t+1|t}\hat{\xi}'_{t+1|t})]J'_t. \end{aligned} \tag{159}$$

The cross-product terms have disappeared from the left side because $\hat{\xi}_{t+1|T}$ is a linear function of $\mathcal{Y}_T$ and so is uncorrelated with the projection error $\xi_t - \hat{\xi}_{t|T}$. Similarly, on the right side, $\hat{\xi}_{t+1|t}$ is uncorrelated with $\xi_t - \hat{\xi}_{t|t}$.

[1] The law of iterated projections states that

$$\hat{E}\xi_t|\mathcal{Y}_t) = \hat{E}[\hat{E}(\xi_t|\mathcal{Y}_T)|\mathcal{Y}_t].$$

The law of iterated projections thus allows us to go from a larger information set to a smaller one.

Equation (159) states that

$$P_{t|T} = P_{t|t} + J_t\{-E[(\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T})] + E[(\hat{\xi}_{t+1|t}\hat{\xi}'_{t+1|t})]\}J'_t. \tag{160}$$

The bracketed term in (160) can be expressed as

$$\begin{aligned}
&- E[(\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T})] + E[(\hat{\xi}_{t+1|t}\hat{\xi}'_{t+1|t})] \\
&= \{E[(\xi_{t+1}\xi'_{t+1})] - E[(\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T})]\} - \{E[(\xi_{t+1}\xi'_{t+1})] - E[(\hat{\xi}_{t+1|t}\hat{\xi}'_{t+1|t})]\} \\
&= \{E[(\xi_{t+1} - \hat{\xi}_{t+1|T})(\xi_{t+1} - \hat{\xi}_{t+1|T})']\} - \{E[(\xi_{t+1} - \hat{\xi}_{t+1|t})(\xi_{t+1} - \hat{\xi}_{t+1|t})']\} \\
&= P_{t+1|T} - P_{t+1|t}.
\end{aligned} \tag{161}$$

The second-to-last equality used the fact that

$$\begin{aligned}
E[\xi_{t+1}\hat{\xi}_{t+1|T}] &= E[(\xi_{t+1} - \hat{\xi}_{t+1|T} + \hat{\xi}_{t+1|T})\hat{\xi}'_{t+1|T}] \\
&= E[(\xi_{t+1} - \hat{\xi}_{t+1|T})\hat{\xi}'_{t+1|T}] + E[\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T}] \\
&= E[\hat{\xi}_{t+1|T}\hat{\xi}'_{t+1|T}],
\end{aligned}$$

since the projection error $(\xi_{t+1} - \hat{\xi}_{t+1|T})$ is uncorrelated with $\hat{\xi}_{t+1|T}$. Similarly, $E(\xi_{t+1}\hat{\xi}_{t+1|t}) = E(\hat{\xi}_{t+1|t}\hat{\xi}'_{t+1|t})$. Substituting (161) into (160) establishes that the smoothed estimate $\hat{\xi}_{t|T}$ has MSE given by

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J'_t. \tag{162}$$

Again, this sequence is generated by moving through the sample backward starting with $t = T - 1$.

2.8.6 Statistical Inference with the Kalman Filter

The calculation of the mean square error

$$P_{\tau|t} = E[(\xi_\tau - \hat{\xi}_{\tau|t})(\xi_\tau - \hat{\xi}_{\tau|t})']$$

described earlier assumed that the parameters of the matrices F, Q, A, H, and R were known with certainty. Section 2.8.4 showed how these parameters could be estimated from the data by maximum likelihood. There would then be some sampling uncertainty about the true values of these parameters, and the calculation of $P_{\tau|t}$ would need to be modified to obtain the true mean squared errors of the smoothed estimates and forecasts.

Suppose the unknown parameters are collected in a vector θ. For any given value of θ, the matrices $F(\theta)$, $Q(\theta)$, $A(\theta)$, $H(\theta)$, and $R(\theta)$ could be used to construct $\hat{\xi}_{\tau|T}(\theta)$ and $P_{\tau|T}(\theta)$ in the formulas presented earlier; for $\tau \leq T$, these are the smoothed estimate and MSE given in (158) and (162), respectively; while for $\tau > T$, these are the forecast and its MSE in (128) and (130). Let $\mathcal{Y}_T \equiv (y'_T, y'_{T-1}, \ldots, y'_1, x'_T, x'_{T-1}, \ldots, x'_1)'$ denote the observed

data, and let θ_0 denote the true value of θ. The earlier derivations assumed that the true value of θ was used to construct $\hat{\xi}_{\tau|T}(\theta_0)$ and $P_{\tau|T}(\theta_0)$.

Recall that the formulas for updating a linear projection and its MSE yield the conditional mean and conditional MSE when applied to Gaussian vectors (see Hamilton, 1994, Chapter 4). Thus, if $\{v_t\}$, $\{w_t\}$, and ξ_1 are truly Gaussian, then the linear projection $\hat{\xi}_{\tau|T}(\theta_0)$ has the interpretation as the expectation of ξ_τ conditional on the data,

$$\hat{\xi}_{\tau|T}(\theta_0) = E(\xi_\tau|\mathcal{Y}_T); \tag{163}$$

while $P_{\tau|T}(\theta_0)$ can be described as the conditional MSE:

$$P_{\tau|T}(\theta_0) = E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)][\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)]'|\mathcal{Y}_T\}. \tag{164}$$

Let $\hat{\theta}$ denote an estimate of θ based on $\mathcal{Y}_T$, and let $\hat{\xi}_{\tau|T}(\theta)$ denote the estimate that results from using $\hat{\theta}$ to construct the smoothed inference or forecast in (158) or (128). The conditional mean squared error of this estimate is

$$\begin{aligned}
&E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\hat{\theta})][\xi_\tau - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\} \\
&\quad = E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0) + \hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})] \\
&\qquad \times [\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0) + \hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\} \\
&\quad = E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)][\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)]'|\mathcal{Y}_T\} \\
&\qquad + E\{[\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})][\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\}.
\end{aligned} \tag{165}$$

Cross-product terms have disappeared from (165), since

$$\begin{aligned}
&E\{[\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})][\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)]'|\mathcal{Y}_T\} \\
&\quad = [\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})] \times E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\theta_0)]'|\mathcal{Y}_T\} \\
&\quad = [\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})] \times 0'.
\end{aligned}$$

The first equality follows because $\hat{\xi}_{\tau|T}(\theta_0)$ and $\hat{\xi}_{\tau|T}(\hat{\theta})$ are known nonstochastic functions of $\mathcal{Y}_T$, and the second equality is implied by (163). Substituting (164) into (165) results in

$$\begin{aligned}
&E\{[\xi_\tau - \hat{\xi}_{\tau|T}(\hat{\theta})][\xi_\tau - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\} \\
&\qquad = P_{\tau|T}(\theta_0) + E\{[\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})][\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\}.
\end{aligned} \tag{166}$$

Equation (166) decomposes the mean squared error into two components. The first component, $P_{\tau|T}(\theta_0)$, might be described as the 'filter uncertainty'. This is the term calculated from the smoothing iteration (162) or forecast MSE (130) and represents uncertainty about ξ_τ that would be present even if the true value θ_0 were known with certainty. The second term in (166),

$$E\{[\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})][\hat{\xi}_{\tau|T}(\theta_0) - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\}$$

might be called 'parameter uncertainty'. It reflects the fact that in a typical sample $\hat{\theta}$ will differ from the true value θ_0.

A simple way to estimate the size of each source of uncertainty is by Monte Carlo integration. Suppose we adopt the Bayesian perspective that θ itself is a random variable. From this perspective, (166) describes the MSE conditional on $\theta = \theta_0$. Suppose that the posterior distribution for the MLE in (139) suggests that $\theta|\mathcal{Y}_T$ might be regarded as approximately distributed $N(\hat{\theta}, (1/T)\mathscr{I}^{-1})$, where $\hat{\theta}$ denotes the MLE. We might then generate a large number of values of θ, say, $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(2000)}$, drawn from a $N(\hat{\theta}, (1/T)\mathscr{I}^{-1})$ distribution. For each draw (j), we could calculate the smoothed estimate or forecast $\hat{\xi}_{\tau|T}(\theta^{(j)})$. The deviations of these estimates across Monte Carlo draws from the estimate $\hat{\xi}_{\tau|T}(\hat{\theta})$ can be used to describe how sensitive the estimate $\hat{\xi}_{\tau|T}(\hat{\theta})$ is to parameter uncertainty about θ :

$$\frac{1}{2000}\sum_{j=1}^{2000}[\hat{\xi}_{\tau|T}(\theta^{(j)}) - \hat{\xi}_{\tau|T}(\hat{\theta})][\hat{\xi}_{\tau|T}(\theta^{(j)}) - \hat{\xi}_{\tau|T}(\hat{\theta})]'. \tag{167}$$

This affords an estimate of

$$E\{[\hat{\xi}_{\tau|T}(\theta) - \hat{\xi}_{\tau|T}(\hat{\theta})][\hat{\xi}_{\tau|T}(\theta) - \hat{\xi}_{\tau|T}(\hat{\theta})]'|\mathcal{Y}_T\},$$

where this expectation is understood to be with respect to the distribution of θ conditional on $\mathcal{Y}_T$.

For each Monte Carlo realization $\theta^{(j)}$, we can also calculate $P_{\tau|T}(\theta^{(j)})$ from (162) or (130). Its average value across Monte Carlo draws,

$$\frac{1}{2000}\sum_{j=1}^{2000}P_{\tau|T}(\theta^{(j)}), \tag{168}$$

provides an estimate of the filter uncertainty in (166),

$$E[P_{\tau|T}(\theta)|\mathcal{Y}_T].$$

Again, this expectation is with respect to the distribution of $\theta|\mathcal{Y}_T$.

The sum of (167) and (168) is then proposed as an MSE for the estimate $\hat{\xi}_{\tau|T}(\hat{\theta})$ around the true value ξ_τ.

3 Data

The dataset consists of 17 quarterly seasonally unadjusted National Accounts' time series and 4 their aggregates, including GDP. The description of the dataset is in Appendix A1.

4 Results

All calculations of the results below are performed in Scilab with the aid of Grocer toolbox. [2]

Direct versus indirect forecasts First, we consider direct versus indirect forecasting, using ad-hoc picked SARIMA models: (011)(011), (010)(011), and (110)(010). The first model is the so-called 'Airline' model discussed in Subsection 2.6.4, frequently used as a benchmark model in seasonal adjustment programs. The second model, (010)(011), is even simpler than the first one and is currently the model used in the seasonal adjustment process of Latvia's GDP. The third model, the AR(1), is usually taken as one of the benchmark models for forecasting comparisons.

All the series are quarterly, starting from $1995Q1$ and ending at $2009Q1$, each with length 57 observations. After one regular and one seasonal differencing, we are left with 52 observations. The regressions are started from 16 observations, so that there are 36 elements per series to compute the Root Mean Squared Forecast Error (RMSE). Thus, for the direct versus indirect forecast comparison, 2016 one-period-ahead real-time forecasts are run whose results are displayed in Table 1. RMSEs are computed from logged series. In order to capture the quality of the forecast at the end of the series, an additional RMSE, named RMSE2ndhalf, is computed for the second half of the series.

There are three kinds of aggregates (see Appendix A1):

- $AO + D21 - D31$;
- $AF + GO + D21 - D31$;
- The sum of 15 single industry time series $+D21 - D31$,

where $D21 - D31$ is net transfers.

Table 1 shows that model (011)(011) gives inferior forecasts than model (010)(011), which is inferior compared to (110)(010). It can also be seen that forecasting the GDP series directly tend to yield better forecasts than any indirect forecasts. The conclusion remains the same when looking at the second half of the forecast period.

Closer look at the models One regular and one seasonal differencing of log GDP yields a series shown in Figure 1.

It is clearly seen that the economic downturn has generated a sequence of negative innovations, rendering the end of the series unstationary. This observation suggests that another regular difference might be needed and that could help improve the precision of forecasts at the end of the series. Applying another regular differencing, we get a series plotted in Figure 2.

[2] Many thanks to Éric Dubois for the maintenance of Grocer.

		(011)(011)	(010)(011)	(110)(010)
B1G	RMSE	0.0397	0.0344	0.0294
	RMSE2ndhalf	0.0529	0.0466	0.0391
AO+D21-D31	RMSE	0.0487	0.0369	0.0321
	RMSE2ndhalf	0.0661	0.0486	0.0431
AF+GO+D21-D31	RMSE	0.0491	0.0386	0.0321
	RMSE2ndhalf	0.0673	0.0502	0.0430
Sum(single)+D21-D31	RMSE	0.0457	0.0369	0.0312
	RMSE2ndhalf	0.0632	0.0493	0.0415

Table 1: Direct versus indirect forecasting.

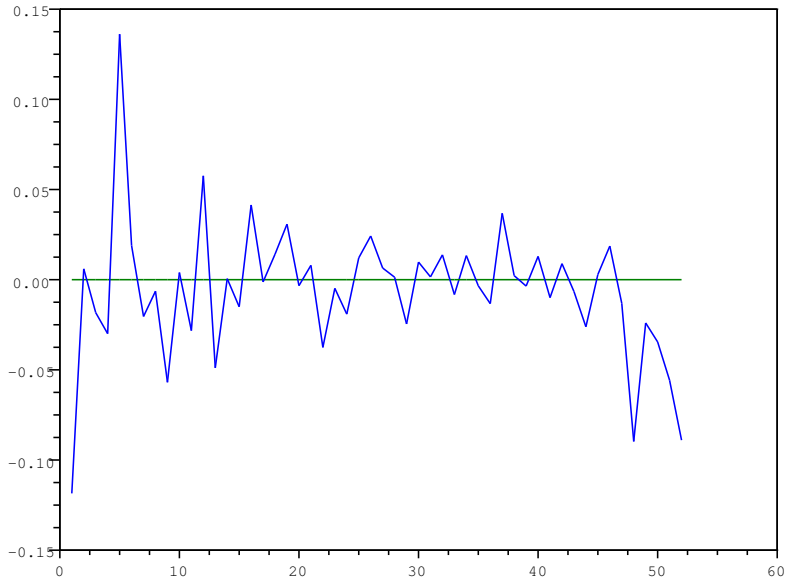

Figure 1: Series after one regular and one seasonal differencing of log GDP.

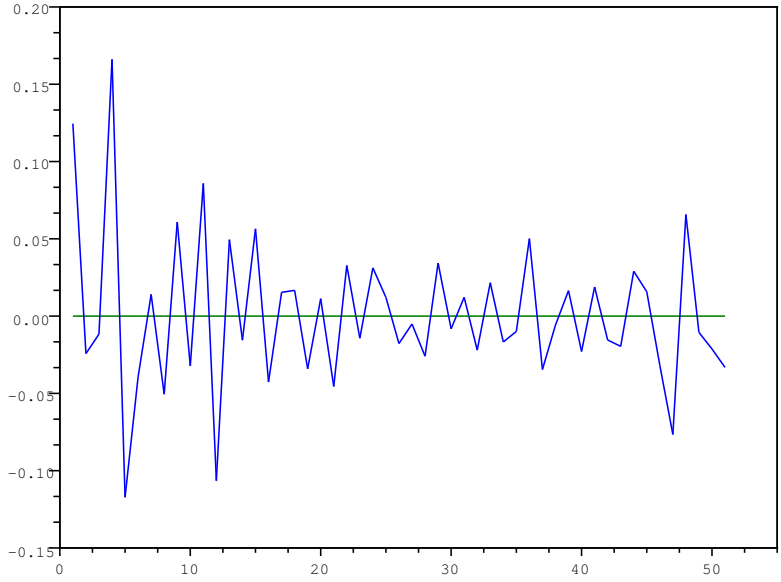

Figure 2: Series after two regular and one seasonal differencing of log GDP.

Now, applying the same three seasonal ARMA models to the new series gives RMSEs shown in Table 2.

The first three models in Table 2 are replicated for a single regular differencing for comparison. Also, since an extra differencing leaves us with one less observation, it is also excluded from the first three models for the RMSE comparison to be fair. Although the series with two regular differences is close to mean zero, we demean it, in line with theory, and present the results with extension "dm". Also, we run a naïve Random Walk (RW) model presented in the last two rows in Table 2.

As expected, an extra regular differencing improves the precision of forecasting at the end of the series, and it also improves the overall RMSE. The rank of the three models stays the same, the $AR(1)$ giving the most precise one-period forecasts. However, there are two surprising observations. First, running models on the demeaned series worsens the quality of forecasts (this remains to be explained). Second, a naïve RW model yields almost as good forecasts as the $AR(1)$ and is the second best of all models.

Model	RMSE	RMSE2ndhalf
(011)(011)	0.0406	0.0529
(010)(011)	0.0347	0.0466
(110)(010)	0.0302	0.0391
(021)(011)	0.0329	0.0436
(020)(011)	0.0302	0.0364
(120)(010)	0.0275	0.0337
(021)(011)dm	0.0367	0.0426
(020)(011)dm	0.0318	0.0375
(120)(010)dm	0.0290	0.0348
naive	0.0293	0.0339
naive dm	0.0303	0.0346

Table 2: A closer look at models.

5 Conclusions

The macroeconomic forecasting literature concentrates on building time-series models for seasonally adjusted series but sometimes the forecasts of seasonally unadjusted series are necessary, in which case we implement seasonal ARIMA models. We have compared the performance of simple seasonal ARIMA models and direct versus indirect real-time one-period forecasting of Latvia's GDP. Four main results are as follows. First,we have seen that the direct forecasting of Latvia's GDP tends to give more precise forecasts than any of the three aggregation levels considered. Second, the AR(1) model seems to perform better, in term of RMSE, than SARMA(01)(01) and SARMA(00)(01). Third, an economic downturn has produced a sequence of negative innovations in a single regularly and single seasonally differenced log GDP, suggesting an extra regular difference might help improve the forecasts at the end of the series. The results show that indeed models SARIMA(021)(011), SARIMA(020)(011), and SARIMA(120)(010) tend to perform better than SARIMA(011)(011), SARIMA(010)(011), and SARIMA(110)(010), respectively. Fourth, a naïve RW model performs very well compared to the rest of the models considered.

References

[1] Ajevskis, V. and Davidsons, G. (2008), "Dynamic Factor Models in Forecasting Latvia's Gross Domestic Product", Working Paper, 2/2008, Latvijas Banka

[2] Anderson, B.D.O., and Moore, J.B. (1979), *Optimal Filtering*, Englewood Cliffs, New Jersey: Prentice-Hall

[3] Barhoumi, K., Darné, O. and Ferrara, L. (2009), "Are disaggregate data useful for factor analysis in forecasting French GDP?", Document de travail No 232, Banque de France

[4] Bell, W.R. and Hillmer, S.C (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models", *Journal of Time Series Analysis* 12, pp.283-300

[5] Benkovskis, K. (2008), "Short-term Forecasts of Latvia's eal Gross Domestic Product Growth Using Monthly Indicators", Working Paper, 5/2008, Latvijas Banka

[6] Boivin, J. and Ng, S. (2006), "Are more data always better for factor analysis?", *Journal of Econometrics* 132, pp.169-194

[7] Box, G.E.P. and Jenkins, G.M. (1970), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day

[8] Brockwell, P. and Davis, R. (1987), *Time Series: Theory and Methods*, Berlin: Springer-Verlag

[9] Burmeister, E., Wall, K.D. and Hamilton, J.D. (1986), "Estimation of Unobserved Expected Monthly Inflation Using Kalman Filtering", *Journal of Business and Economic Statistics* 4, pp.147-60

[10] Caggiano, G., Kapetanios, G. and Labhard, V. (2009), "Are more data always better for factor analysis? Results for the Euro Area, the sex largest Euro Area countries and the UK", Working Paper Series No 1051/May 2009, European Central Bank

[11] Caines, P.E. (1988), *Linear Stochastic Systems*, New York: Wiley

[12] Casals, J., Jerez, M. and Sotoca, S. (2000), "Exact Smoothing for Stationary and Nonstationary Time Series," *International Journal of Forecasting*, vol. 16, pp.59-69

[13] Chen, C. and Liu, L.M. (1993), "Joint Estimation of Model Parameters and Outlier Effects in Time Series", *Journal of the American Statistical Association* 88, pp.284-297

[14] De Jong, P. (1991), "The Diffuse Kalman Filter", *Annals of Statistics* 19, pp.1073-1083

[15] Dempster, A.P., Laird, N. M. and Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society* Series B, 39, pp.1-38

[16] Diebold, F.X. and Mariano, R.S. (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics* 13, pp. 253-365

[17] Dreger, C. and Schumacher, C. (2002), "Estimating large-scale factor models for economic activity in Germany: Do they outperform simpler models?", HWWA Discussion Paper 199, Hamburg Institute of International Economics

[18] Eickmeier, S. and Ng, T. (2009), "Forecasting national activity using lots of international predictors: an application to New Zealand", Discussion Paper, Series 1: Economic Studies, No 11/2009, Deutsche Bundesbank

[19] Fischer, B. and Planas, C. (1998), "Large Scale Fitting of ARIMA Models and Stylized Facts of Economic Time Series", Eurostat Working Paper 9/1998/A/8

[20] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2003), "The Generalized Dynamic Factor Model, One-sided Estimation and Forecasting", LEM Working Paper Series, 2003/13

[21] Gevers, M. and Wertz, V. (1984), "Uniquely Identifiable State-Space and ARMA Parameterization for Multivariable Linear Systems", *Automatica* 20, pp.333-47

[22] Ghosh, D. (1989), "Maximum Likelihood Estimation of the Dynamic Shock-Error Model", *Journal of Econometrics* 41, pp.121-43

[23] Gourieroux, C. and Monfort, A. (1990), *Séries Temporelles et Modéles Dynamiques*, Paris: Economica

[24] Gómez, V. and Maravall, A. (2000a), "Automatic Modelling Methods for Univariate Series", Working Paper 9808, Research Department, Bank of Spain

[25] Gómez, V. and Maravall, A. (2000b), "Seasonal Adjustment and Signal Extraction in Economic Time Series", Working Paper 9809, Research Department,Bank of Spain

[26] Gómez, V. and Maravall, A. (1996), "Programs TRAMO and SEATS; Instruction for the User", Working Paper 9628, Research Department, Bank of Spain

[27] Gómez, V. and Maravall, A. (1994), "Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman filter", *Journal of American Statistical Association* 89, pp.611-624

[28] Hamilton, J.D. (1985), "Uncovering Financial Market Expectations of Inflation", *Journal of Political Economy* 93, pp.1224-41

[29] Hamilton, J.D. (1994): *Time Series Analysis*, Princeton, New Yersey: Princeton University Press

[30] Hannan, E.J. (1971), "The Identification Problem for Multiple Equation Systems with Moving Average Errors", *Econometrica* 39, pp.751-65

[31] Harvey, A.C. (1989), *Forecasting Structural Time Series and the Kalman Filter*, Cambridge: Cambridge University Press

[32] Harvey, A. and Phillips, G.D.A (1979), "Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances", *Biometrika* 66, pp.35-45

[33] Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press

[34] Hillmer, S.C., Bell, W.R. and Tiao, G.C. (1983), "Modelling Considerations in the Seasonal Adjustment of Economic Time Series", in Zellner, A. (ed.), *Applied Time Series Analysis of Economic Data*, Washington, D.C.: U.S. Department of Commerce-Bureau of the Census, pp.74-100

[35] Hylleberg, S., Engle, R.F., Granger, C.W.J. and Yoo, B.S (1990), "Seasonal Integration and Cointegration", *Journal of Econometrics* 44, pp.215-238

[36] Kaiser, R. and Maravall, A. (2001), *Measuring Business Cycles in Economic Time Series*, New York: Springer-Verlag

[37] Kalman, R.E. (1960), "A New Approach to Linear Filtering and Prediction problems", *Journal of Basic Engineering, Transactions of the ASME* Series D, 82, pp.35-45

[38] Kohn, R. and Ansley, C.F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data", *Journal of the American Statistical Association* 81, pp.751-761

[39] Kuzin, V., Marcellino, M. and Schumacher, C. (2009), "MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area", Discussion Paper, Series 1: Economic Studies, No 07/2009, Deutsche Bundesbank

[40] Kuzin, V., Marcellino, M. and Schumacher, C. (2009), "Pooling versus model selction for nowcasting with many predictors: an application to German GDP", Discussion Paper, Series 1: Economic Studies, No 03/2009, Deutsche Bundesbank

[41] Maravall, A. (2000), "An application of TRAMO and SEATS", Working Paper 9914, Research Department, Bank of Spain

[42] Masten, I., Benerjee, A. and Marcellino, M. (2009), "Forecasting with factor-augmented error correction models", EUI Working Papers, RSCAS 2009/32

[43] Morf, M., Sidhu, G.S., and Kailath, T. (1974), "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete-Time Systems", *IEEE Transactions on Automatic Control*, AC 19, pp.315-323

[44] Newbold, P. (1983), "Model Checking in Time Series Analysis", in A. Zellner (ed.) *Applied Time Series Analysis of Economic Data*, Washington, D.C.: U.S. Department of Commerce-Bureau of the Census, pp.133-143

[45] Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients", *Journal of Econometrics* 13, pp.341-63

[46] Rothenberg, T.J. (1971), "Identification in Parametric Models", *Econometrica* 39, pp.577-91

[47] Schumacher, C. (2009), "Factor forecasting using international targeted predictors: the case of German GDP", Discussion Paper, Series 1: Economic Studies, No 10/2009, Deutsche Bundesbank

[48] Shumway, R.H. and Stoffer, D.S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm", *Journal of Time Series Analysis* 3, pp.253-64

[49] Stock, J.H., and Watson, M.W. (2004), "Forecasting with many predictors", Prepared for the *Handbook of Economic Forecasting*

[50] Stock, J.H., and Watson, M.W. (2003), "Combination Forecasts of Output Growth", Working Paper

[51] Stock, J.H., and Watson, M.W. (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, April 2002, Vol. 20, No. 2

[52] Stock, J.H., and Watson, M.W. (1998), "Diffusion Indexes", NBER Working Paper Series, w6702

[53] Stock, J.H., and Watson, M.W. (1991), "A Probability Model of the Coincident Economic Indicators", in Lahiri, K. and Moore, G.H. (eds.) *Leading Economic Indicators: New Approaches and Forecasting Records*, Cambridge, England: Cambridge University Press

[54] Tiao, G.C. and Tsay, R.S. (1983), "Cosistency Properties of Least Squares Estimates of Autoregressive Parameters in ARMA Models", *The Annals of Statistics* 11, pp.856-871

[55] Tiao, G.C. and Tsay, R.S. (1989), "Model Specification in Multivariate Time Series", *Journal of the Royal Statistical Society* B, 51,pp.132-141

[56] Wall, K.D. (1987), "Identification Theory for Varying Coefficient Regression Models", *Journal of Time Series Analysis* 8, pp.359-71

[57] Wang, M. (2008), "Comparing the DSGE model with the factor model: an out-of-sample forecasting experiment", Discussion Paper, Series 1: Economic Studies, No 04/2008, Deutsche Bundesbank

[58] Watson, M.W. (1989), "Recursive Solution Methods for Dynamic Linear Rational Expectations Models", *Journal of Econometrics* 41, pp.65-89

[59] Watson, M.W. and Engle, R.F. (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC, and Varying Coefficient Regression Models", *Journal of Econometrics* 23, pp. 385-400

APPENDIX A1

The list of the National Accounts' time series used in the paper. All series are chained priced as of 2000.

A: Agriculture, hunting and forestry

B: Fishing

C: Mining and quarrying

D: Manufacturing

E: Electricity, gas and water supply

F: Construction

G: Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods

H: Hotels and restaurants

I: Transport, storage and communication

J: Financial intermediation

K: Real estate, renting and business intermediation

L: Public administration and defense; compulsory social security

M: Education

N: Health and social work

O: other community, social and personal service activities

D21: Taxes

D31: Subsidies

AF: The aggregate of A to F

GO: The aggregate of G to O

AO: The aggregate of A to O

B1G: The Gross Domestic Product